DOCUMENT RESUME

ED 445 330 CS 014 134

AUTHOR	Eriksson, Asa
TITLE	Thinking Forwards and Backwards: Metamemory and Metacomprehension Abilities and Strategies in Text Processing.
INSTITUTION	Linkoping Univ. (Sweden). Dept. of Education and Psychology.
ISBN	ISBN-91-7219-838-9
ISSN	ISSN-1102-7517
PUB DATE	2000-00-00
NOTE	149p.; Linkoping Studies in Education and Psychology Dissertation No. 70. Attached to this document are 3 documents on which the present document was based: reprints of 2 journal articles (which are not made available through ERIC due to copyright restrictions); and a manuscript by the author.
PUB TYPE	Collected Works - General (020) -- Information Analyses (070)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	Foreign Countries; *High School Students; High Schools; *Metacognition; *Reading Comprehension; *Reading Processes; Reading Research; *Recall (Psychology)

ABSTRACT

The aim of the present thesis was to investigate high-school students' metamemory and metacomprehension of texts. In three studies, the students read texts and then made prospective as well as retrospective ratings of their own immediate and delayed performance (i.e., measured via text recall and answering performance of comprehension questions). The data have been viewed overall and for different verbal skill groups. Different types of instructions, time of test, placement of rating, types of texts and characteristics of texts have been used. The overall pattern of data suggests that the students accurately predicted and postdicted their text recall. Delayed postdiction accuracy was found, even after a long delay. The pattern for comprehension was not as straightforward, in the sense that the studies demonstrated different results regarding calibration accuracy. However, the students postcalibrated more accurately their comprehension. From a verbal skill perspective, high performing students excelled in performance but the low performing students made the most accurate ratings of memory performance. Irrespective of verbal skill, the students demonstrated study preferences for both memory and comprehension of texts. These preferences interacted with text recall but not with answering performance on the comprehension questions. The results suggest that effort is a key concept to consider in this line of research. First, the students found reading to remember a more effort requiring task than reading to comprehend. This supposedly resulted in better awareness of memory performance than comprehension of the same texts. Also, the reading instruction that emphasizes learning yielded both immediate and delayed prediction accuracy. This instruction was regarded as requiring the most effort. Second, the better the person's verbal ability, the less attention he or she requires to complete the reading task, with the best possible outcome as a result. High verbal skill reading is presumably effortless and automatized. Third, when students studied texts in their most preferred way it again resulted in best possible text recall, but reduced prediction accuracies. Taken together, metacognitive thinking seems to be

Reproductions supplied by EDRS are the best that can be made
from the original document.

most useful in the beginning of and in the development of skill. Contains 109 references and 7 tables of data. Appended is a related manuscript by the author. (Author/RS)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 445 330

THINKING FORWARDS AND BACKWARDS

Metamemory and Metacomprehension Abilities and Strategies in Text Processing

Åsa Eriksson

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

FACULTY OF ARTS AND SCIENCE
LINKÖPINGS UNIVERSITET

Linköping Studies in Education and Psychology No. 70
Linköpings universitet, Department of Behavioural Sciences

Linköping 2000

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
C. Holmberg
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

CS 014 134

BEST COPY AVAILABLE

Linköping Studies in Education and Psychology Dissertation No. 70

THINKING FORWARDS AND BACKWARDS

Metamemory and Metacomprehension Abilities and Strategies in Text Processing

Åsa Eriksson

Akademisk avhandling

Som med vederbörligt tillstånd av filosofiska fakulteten vid Linköpings universitet för avläggande av filosofie doktorsexamen kommer att offentligt försvaras på institutionen för beteendevetenskap, Eklundska salen (I:101), fredagen den 6 oktober 2000, kl. 13.00

Abstract

The aim of the present thesis was to investigate high-school students' metamemory and metacomprehension of texts. In three studies, the students read texts and then made prospective as well as retrospective ratings of their own immediate and delayed performance (i.e., measured via text recall and answering performance of comprehension questions). The data have been viewed overall and for different verbal skill groups. Different types of instructions, time of test, placement of rating, types of texts and characteristics of texts have been used. The overall pattern of data suggests that the students accurately predicted and postdicted their text recall. Delayed postdiction accuracy was found, even after a long delay. The pattern for comprehension was not as straightforward, in the sense that the studies demonstrated different results regarding calibration accuracy. However, the students postcalibrated more accurately their comprehension. From a verbal skill perspective, high performing students excelled in performance but the low performing students made the most accurate ratings of memory performance. Irrespective of verbal skill, the students demonstrated study preferences for both memory and comprehension of texts. These preferences interacted with text recall but not with answering performance on the comprehension questions. The results suggest that effort is a key concept to consider in this line of research. First, the students found reading to remember a more effort requiring task than reading to comprehend. This supposedly resulted in better awareness of memory performance than comprehension of the same texts. Also, the reading instruction that emphasizes learning, yielded both immediate and delayed prediction accuracy. This instruction was regarded as the most effort requiring. Second, the better the person's verbal ability, the less attention he or she requires to complete the reading task, with the best possible outcome as a result. High verbal skill reading is presumably effortless and automatized. Third, when students studied texts in their most preferred way it again resulted in best possible text recall, but reduced prediction accuracies. Taken together, metacognitive thinking seems to be most useful in the beginning of and in the development of skill.

Keywords: Metacognition, metamemory, metacomprehension, predictions, calibrations, text comprehension, text recall, verbal skill, effort, study preferences.

Department of Behavioural Sciences
Linköpings universitet, S-581 83 Linköping, Sweden
Linköping 2000

ISRN LiU-IBV-STU- -70 ISBN 91-7219-838-9 ISSN 1102-7517

THINKING FORWARDS AND BACKWARDS

Metamemory and Metacomprehension Abilities and Strategies in Text Processing

Åsa Eriksson

FACULTY OF ARTS AND SCIENCE
LINKÖPINGS UNIVERSITET

Linköping Studies in Education and Psychology No. 70
Linköpings universitet, Department of Behavioural Sciences

Linköping 2000

LINKÖPINGS UNIVERSITET
Department of Behavioural Sciences
SE-581 83 Linköping

Thinking Forwards and Backwards
Metamemory and Metacomprehension
Abilities and Strategies in Text Processing

Åsa Eriksson

Tryck:UniTryck, Linköping 2000
ISRN LiU-IPP-STU--70--SE
ISBN 91-7219-838-9
ISSN 1102-7517

PREFACE

This thesis in psychology is based on the present summary and the following studies.

(I) Gillström, Å & Rönnberg, J. (1994). Prediction accuracy of text recall: ease, effort and familiarity. *Scandinavian Journal of Psychology, 35*, 367-386.

(II) Gillström, Å & Rönnberg, J. (1995). Comprehension calibration and recall prediction accuracy of texts: reading skill, reading strategies, and effort. *Journal of Educational Psychology, 87*, 545-558.

(III) Eriksson, Å. (2000). Comprehension calibration and memory prediction accuracy of texts: the effect of reading to remember and reading to comprehend instructions on performance and accuracy of ratings. Submitted manuscript.

Acknowledgement

Some are fortunate to have great colleagues and a social network to hang on to in life, and I am one of those. The time has come to thank you all for your patience and support.

To begin with, this thesis would not have come around without you, Jerker. As a supervisor you have kept me going, to think again and to better express myself. It has been tough, but at the same time, challenging to try and meet up to your standards. I have learned a lot during the years. Also, having a whole ”gang of boys" to take care of, I have wanted to give up at times, but you have encouraged me to continue. Thanks!

Stefan S and Stefan G thank you for valuable comments and suggestions to the summary. During the years you always have taken the time to listen and suggest new approaches and hypotheses - something that you both are good at! Björn, Janne and Ulf , thank you for sharing an interest in psychological research and for inspiring ideas. This holds true for the whole CCDD group as well.

Karina´s ”room and board" has had a very crucial influence on this thesis. I do not think we agree on anything, Karina, which makes our relation so special. Thank you for being a good friend, for all our sandwiches and never-ending discussions!

Ulla-Britt, you have been an excellent proof-reader and an interested and interesting colleague to discuss metacognition with. Birgitta R, Kerstin J and Titti, thank you for "administering" me in the best possible way.

I wish to express a special thanks to my mother and father, Lena and Arne. You believe in me as nobody else and help out in any way you can. My mother in law, Lonny, is also the best. You take care of the children anytime it is needed. Being a mother of three, I can not thank you all enough! I also wish to thank Ingrid, Anette and Eva and their families for being good friends and always supportive.

Finally, Mattias, Fredrik and Ludvig, you are the meaning of my life. I know I have been especially busy the last couple of weeks. Payback time has come, though, let's go to the ”badhus" every week, play ball and do other fun things together!

Åsa Eriksson, September 2000

Contents

1. Introduction 9
    - 1.1 Knowledge of cognition and regulation of cognition 11
    - 1.2 Areas of metacognitive research 12
    - 1.3 Early contributions to metacognition 14
    - 1.4 Current aspects of metacognitive models 16
    - 1.5 Self-knowledge and reflection 17
2. Metacognitive development 18
3. Definition of metacomprehension 22
4. Definitions of metamemory 23
5. Metacomprehension and metamemory combined 26
6. Factors being studied 28
    - 6.1 Verbal ability 29
    - 6.2 Text processing 31
    - 6.3 Time of test 32
    - 6.4 Prospective and retrospective ratings 33
    - 6.5 Type of text material 34
7. Design and purpose 35
    - 7.1 Overall design 36
    - 7.2 Open-ended questions 39
8. Description of Studies I-III 39
    - 8.1 Study I 39
    - 8.2 Study II 45
    - 8.3 Study III 48
    - 8.4 Summary of data patterns 49
9. Discussion 50
    - 9.1 Metamemory and metacomprehension results 51
    - 9.2 Additional data analyses and validity aspects 56
    - 9.3 What' s new - combining metacomprehension and metamemory 58
    - 9.4 What's new - verbal skill 60
    - 9.5 What' s new - studying preferences 62
    - 9.6 Three critical points 63
    - 9.7 Main conclusions 64
    - 9.8 Future directions 66
10. References 68

1. Introduction

Metacognition is about how we manage to think about our own thoughts. A person's thoughts can revolve around what he or she knows, what he or she currently is doing or what his or her current cognitive state is – labeled metacognitive knowledge, metacognitive skill, and metacognitive experience, respectively. These conceptions are based on a person's internal mental representations of reality and how he or she manages to appreciate and evaluate these representations (Dunlosky, 1998). Hacker (1998) suggests that

> *"there does seem to be general consensus that a definition of metacognition should include at least these notions: knowledge of one's knowledge, processes, and cognitive and affective states; and the ability to consciously and deliberately monitor and regulate one's knowledge, processes, and cognitive and affective states*" (page 11).

Lories, Dardenne and Yzerbyt (1998) propose that metacognition is one of the fundamental characteristics of human cognition in that we have the ability to think about our own cognitive acts. For instance, after having read a text we can evaluate our comprehension and, if necessary, for example look up the meaning of words. In this sense, text reading per se is the cognitive act whereas thoughts embedding the reading task are denoted metacognitive. By inference, metacognition is concerned with how cognitive acts apply to themselves. Davidson and Sternberg (1998) suggest that one hallmark of metacognitive ability is a correct transfer of a strategy from one problem to another, that is, knowing when and where to use a certain strategy.

Koriat (1998) contends that cognitive acts are often accompanied by metacognitive operations. Taking a test could include knowing the answers but also more or less strong senses of feeling-of-knowing the answers. In a problem-solving situation, a person has to consider what to do, what is needed and if he or she knows how to approach the task effectively.

These considerations are subjective – metacognitive, and have as such a measurable effect on our behavior. For example, if someone has to construe something he or she might have to use a description. If this person is familiar with such descriptions, he or she might feel confident and at ease. If not, he or she might expect trouble. Thoughts like these could affect how you subsequently approach and manage the task (Koriat, 1998, Koriat & Goldsmith, 1997).

This thesis departs from the fact that we, as human beings, have the ability to consciously and deliberately plan, monitor, evaluate, and, if necessary, improve cognitive actions (Dunlosky, 1998; Hacker, 1998; Lories, et, al., 1998; Koriat, 1998). The overall purpose has been to study factors related to students' cognitive monitoring of their reading comprehension and memory of text (Hacker, 1998; Koriat, 1998). Already in 1979, Flavell presented a model describing metacognition as the knowledge we have of our own cognitive processes and behavior (Flavell, 1979). The foundation of this model is that people know how to reflect, they are aware of their own reflections, and possess valuable knowledge about cognition. In other words, learners possess metacognitive knowledge about person, task and strategy variables (Flavell, 1979; Garner, 1987; Lin & Zabrucky, 1998). In each of the present studies, students rated how well they had comprehended the text and how much of the text they would be able to recall. This thesis have included the person, task and strategy variables in that overall and verbal skill analyses have been made (person), different types of texts and instructions have been used, the students have made both prospective and retrospective ratings, and finally, on-line as well as long-term metacognition have been investigated (task, strategy).

A thesis about cognitive monitoring includes both psychological and educational implications. It is obvious that some students learn easily, whereas others have to struggle hard. But, what kind of knowledge do these groups of students have about their own performance? Do they differ or think alike? At the time of data collection for this thesis, metacognition had become an interesting object of study, addressing these and similar questions (Garner, 1987; Pramling, 1987). Ever since, metacognitive research has been conducted in many different ways and from different perspectives (e.g., Hacker, 1998; Persson, 1994). The present thesis has had a quantitative focus in which both cognitive and metacognitive data have been collected. Objective and subjective measures of reading performance have been analyzed. In some of the studies, open-ended questions have been collected regarding students' views of the experiment and instructions. This qualitative aspect of data collection was mainly added to broaden the main set of data.

The first part of the thesis will introduce metacognition as a concept, different areas of research, and results. It will continue to describe two concepts that underlie metacognition - metacomprehension and metamemory – which are the focus of the present thesis.

1.1 Knowledge of cognition and regulation of cognition

Lin and Zabrucky (1998) discuss two different aspects of metacognition, one of which is concerned with knowledge of cognition and the other with regulation of cognition (cf. Baker & Brown, 1984). Pramling (1987) indicates that both of these are closely related and supportive of each other. In a reading situation, people typically possess knowledge indicating that certain types of texts or contents are easier to understand than others. This type of knowledge is regarded as rather stable and something that you will not suddenly forget. It does however require that the learner can view his or her cognitive processes as an object for reflection and thought. Thus, it is a skill that gradually develops both in content and maturity (Pramling, 1987).

The regulation of cognition suggests that people also possess knowledge regarding what type of strategy to use in order to understand a certain text successfully. This type of knowledge is not as stable. If you have a lot on your mind you might not use strategies as effectively as you otherwise would (Lin & Zabrucky, 1998). It also seems that this knowledge is sensitive towards task and context conditions. That is, a person who shows metacognitive skills in writing does not automatically show the same skills in reading (Pramling, 1987). Bouffard, Boisvert, Vezeau and Larouche (1995) suggest that there are three major components of successful self-regulation. The first component refers to the cognitive strategies we have to learn and know about, such as memorizing and understanding. The second to the metacognitive strategies we need to know for adequate supervision during task execution. The third component refers to amount of motivation needed to solve the tasks.

1.2 Areas of metacognitive research

Hacker (1998) proposes four main metacognitive areas of investigation. The first is concerned with cognitive monitoring. The second with regulation of thinking processes to cope with changes. The third is concerned with a combination of the first and second type of studies. The fourth area is concerned with practical educational aspects of metacognition.

The present thesis is concerned with the first of these categories - cognitive monitoring. This area of research gives information pertinent to whether students can identify what they know and do not know, what they have and have not learned, and also if they can use this knowledge effectively. One phenomenon is the *tip-of-the-tongue,* which is quite common experience of having the answer literally on the tip of the tongue but being unable to produce it (Sinkavich, 1995). According to Baddeley (1999), a person who claims that he or she knows the answer is there is usually correct, given the appropriate prompt. Another and similar phenomenon is the *feeling-of-knowing* (FOK) which is tested for example by a person answering questions like *"Who was the first person who walked on the moon?"*. To all questions answered incorrectly, the subjects rate whether or not they can point out the right answer from given alternative answers. Carroll and Nelson (1993) concluded that these FOK's are quite valid indicators of the contents of a person's memory. Based on a person' s domain knowledge, *ease of learning* ratings taps into how difficult the student feel it would be to learn new information from this domain. During or at the end of learning students can make *judgements of learning* which indicate how likely it is that the students will remember a studied item, that is, whether it has been learned or not (Carroll, Nelson, & Kirwan, 1997). The present thesis is concerned with *performance predictions* (in this thesis labeled predictions) which means that students rate how well they will do on a future test.

The second metacognitive areas of investigation is concerned with students' ability to use a learned strategy in a new situation (Hacker, 1998). This line of research investigates whether or not students are able to change or alter behavior to match new perspectives. Interest is also placed on the strategy itself. Early on, mentally retarded people were quite often used as subjects. Today these studies usually include a training and strategy transfer task. Hacker (1998) concludes that enhancing students' metacognitive awareness of the usefulness and function of a strategy usually improves learning, effective strategy use and thus, subsequent transfer (Bristow, Cowley & Davies, 1998).

To third area of metacognitive investigate study both cognitive monitoring and the way students are able to learn new strategies and transfer them into new domains is sometimes labeled *"metacognition in action"* (Hacker, 1998). One typical type of task is sort recall which asks students to recall as many items as possible. They are given a list of words and are then required to monitor their processing of the list and also to use different strategies that will improve the amount of recalled items.

While the first decades of metacognitive research has been concerned with theory-building, there is at present, a growing interest in practical metacognition. Thus, the fourth category investigate the usefulness of improving and informing students of metacognition, how to solve problems, how they think during reading, and so forth (Hacker, 1998). As a consequence, there is a growing interest in educational application. Bristow, et al. (1998), suggest different ways to improve young children's learning, for instance through better personal organization. In different ways, memory awareness can be raised and if children are taught effective strategies it can lead to better storage and retrieval of information. For children with poor memory improvement of personal organization can be achieved through daily routines, short and to-the-point instructions, using external memories (e.g., diaries, timetables) and visualizing techniques.

1.3 Early contributions to metacognition

Flavell (1979) was one of the first modern contributors to metacognition. He suggested that metacognition refers to cognition about cognition. Among other things, cognition is concerned with comprehension and memory, and in this sense metacognition is concerned with thinking about comprehension and memory. Metacognition could be viewed as an umbrella concept with metamemory and metacomprehension underlying this superordinate term (Garner, 1987). Flavell (1979) divided metacognition into metacognitive knowledge, metacognitive experience and strategy use.

Metacognitive knowledge is concerned with three different variables covering people's world knowledge and what they know about their own cognitive abilities, goals, and actions. The person variable, refers to the knowledge people have of their own nature or the nature of another person. This could include intraindividual (e.g., that I am, as a person, better at doing one type of task than with another type of task), inter-individual (e.g., how I do in comparison with my peers) or universal knowledge (e.g., that certain materials need more careful consideration than others) (Flavell, 1979; Garner, 1987). Secondly, the task variable, which refers to the requirement of the task and how to meet these requirements. This could include the amount of available information a person has when a task is being solving. If it is something you are familiar with you might act in a certain way as opposed to if the information is unfamiliar. Third, strategies, which refer to the ways and methods people use to reach a goal (Hacker, 1998). The third variable could include knowing when and how to apply a certain strategy. Quite often two or all three of these variables are combined and/or interact with each other (Flavell, 1979; Garner, 1987). A fourth category, suggested by Lin and Zabrucky (1998), is the text variable which investigates how text manipulations affect metacognitive monitoring. The present thesis have investigated different aspects of these four variables and their effect on metacomprehension and metacognition. Garner (1987) pointed out that metacognitive knowledge is similar to other types of knowledge in that it could be declarative as well as procedural. As metacognitive knowledge gradually improves with experience, it could be activated more or less automatically.

Metacognitive experience refers to any cognitive or affective experience accompanying an intellectual task. This could include knowing that you do not understand, to use experiences from solving one task to solving another, or the feeling of success or failure (Hacker, 1998). These experiences can happen before (e.g., personal strength), during (e.g., strategy knowledge) or after (e.g., task information) a cognitive act (Flavell, 1979: Garner, 1987). One typical metacognitive experience is the earlier mentioned "tip-of-the-tongue feeling" that can pop up in the mind of a person as he/she feels that they know the answer but fail to recall it (Garner, 1987). Also, learners might have feelings of confusion when they fail to solve a task. To elicit metacognitive experience you have to ask questions like, *"Do I understand?"* (Garner, 1987). According to Flavell (1979) metacognitive knowledge and metacognitive experience do not differ in quality, only in content and function. Metacognitive knowledge can lead to different metacognitive experiences. For example, it is one thing to be familiar with a task, quite another to be unfamiliar (Hacker, 1998).

The *strategy use* is concerned with the actions a person makes to reach a goal. Sometimes the metacognitive experience can affect this latter two via the establishment of new goals or revision of old ones. A look at the three aspects of metacognition suggests that metacognitive knowledge is the base for metacognitive experience and choice of strategy to reach a specific goal. Then again, metacognitive experience can also alter metacognitive knowledge, and so forth (Garner, 1987).

1.4 Current aspects of metacognitive models

Bouffard et al. (1995) argued that modern metacognitive models often describe three major components of self-regulation: cognitive strategies, such as memorizing, metacognitive strategies, for example supervising ongoing problem solving, and finally, motivation. Some students seem to be more strategically involved in their cognitive endeavors. They plan each step, they evaluate along the way and try to choose the best possible strategy that a certain task requires. Other students seem to be solving tasks more arbitrarily, maybe because they are told to act in a certain way. Earlier models of metacognition did not stress the need of motivational and affective variables. According to Bouffard, Vezeau and Bordeleau (1998), both of these aspects are important factors to consider in order to understand how metacognition develops and also to identify when a person is likely to act metacognitively. Bouffard, et al. (1995) argued that motivated students are more likely to engage in the rather effortful and time-consuming strategic behavior that is typical of metacognition (cf. Bouffard, Markovits, Vezau, Boisvert & Dumas, 1998). Baddeley (1999) claimed that there exists a connection between motivation to learn and amount of time and attention a person spends on a material. Thus, due to the least effort-principle, it seems as if successful learning requires knowing how to learn but also the motivation to do so (Graham, Harris, & Troia, 1998).

To sum up, the present thesis is concerned with cognitive monitoring which is one of four areas of metacognitive investigation (Hacker, 1998). Research on cognitive monitoring potentially offers information about people's ability to evaluate their own cognitive performance. Flavell (1979) presented a model of metacognition which included *metacognitive knowledge*, *metacognitive experience*, and *strategy use*. The first refers to people's knowledge about themselves and others. The second to feelings accompanying cognitive tasks. The third to people's actions to reach the perception of the goal. Current models also include motivation as an important prerequisite for metacognition and effective learning (Baddeley, 1999; Bouffard, et al., 1995).

1.5 Self-knowledge and reflection

It is important that a person knows about his or her self, and it has been found that high levels of self-esteem have a positive effect on performance (Bouffard, 1998; Davies & Brember, 1999). In addition, metacognition requires reflection, that is thinking about your own thoughts, and, if necessary, action related to them. Lee and Hutchinson (1998) presented different ways to improve students' learning processes via reflection. One way was to add questions that made students reflect on what they have just read or are supposed to learn. In their study, students with low knowledge skill gained the most from these questions. To improve learning, obviously the questions need to be "right", asked at the right time and given to the students after having read the text. Another way to improve learning via reflection is elaboration to clarify the text (Lee & Hutchinson, 1998).

Davidson and Sternberg (1998) reported on training programs that were used to improve metacognitive knowledge. These programs used think-aloud situations, study techniques and question guide-lines. The purpose of these programs was to find out what students know about different strategies, how they use them, and also to find out if they know when and where to use them. The current trend seems to be projected towards reciprocal teaching in which the corner-stones are social interactions, guided questioning and applicability across different domains (Davidson & Sternberg, 1998).

Due to the least-effort principle students do not always take necessary actions to improve their learning (Graham, et al., 1998). To solve a problem, a person needs to be aware of and able to manage his or her mental activities to consider the givens, goals and obstacles of a problem (Davidson & Sternberg, 1998). Givens are those conditions that form the initial problem. Davidson and Sternberg (1998) claimed that poor encoding of a problem could be the result of poor metacognitive knowledge about procedures. A problem-solver has to create internal pictures of the givens and of the relation among givens. There is not one single perfect representation but different ones depending on matters such as cognitive abilities and styles. Once a person has encoded the problem he or she has to plan how to reach the goal which often requires problem decomposition. The division into subgoals usually results in fewer errors compared to global solutions. The three metacomponents involved in effective planning are selecting the strategic components to use on the problem, sequencing these in a facilitative way, and finally allocating attention. Quite often planning relies on heuristics such as means-end analysis (Davidson & Sternberg, 1998). Typical obstacles that stand in the way of a solution are stereotypy, and lack of plans or procedures due to novelty or unfamiliarity with the problem. Sometimes the problem-solver does not monitor or evaluate ongoing processes. It is important that the learner keeps track on past, present and future activities and how close to a solution the he or she is.

This first section of the thesis has introduced the concept of metacognition, presented different areas of research, and given some ideas as to why metacognition is important to study. This thesis departs from the assumption that we are able to think about our own thoughts. The next section will describe different developmental aspects of metacognition.

2. Metacognitive development

Developmental researchers have shown a great deal of interest in metacognition as they have studied how children mature cognitively as well as in their ability to think about cognition. It has also been important to find out how children learn to appreciate themselves and their own abilities (Nelson et al., 1998). These latter aspects are important, since the self-system regulates behavior and motivates actions. A positive self-system requires self-esteem, perceived competence, self-efficacy and control of success and failure (Bouffard, 1998). According to Zimmerman (2000), educators are well aware of the relationship between learners' beliefs about their own academic capabilities and their motivation to achieve.

Davies and Brember (1999) suggest that there exists a positive correlation between self-esteem and successful individual functions. Self-concept is an umbrella term which consists of terms like self-image, ideal self, and self-esteem. The first aspect is concerned with what a person is, the second with what a person would like to be and the third with the discrepancy between the first and the second. In this sense, high self-esteem occurs when there is agreement between what a person is and what the person would like to be (Davies & Brember, 1999). The way the self-system develops depends on several factors, such as individual characteristics, social comparison, school environment and parents' attitudes (Bouffard, 1998). Davies and Brember (1999) indicated that self-esteem usually declines between the ages 6 and 12. At this point in life children realize that they can not always live up to others' expectations, which reduces their self-esteem. In early adulthood self-esteem usually increases again.

According to Bouffard et al. (1998), self-perception of competence affects how people commit to different tasks, how much effort they invest, and also their self-regulation of learning. There is a clear age-related pattern, such that the self-perception of children in kindergarten and first grade is more unrealistic. Children become more self-aware as they grow older. To appraise competence, a child has to be able to evaluate, weigh and compare past and present experiences. This ability to reflect on your cognitive status gradually develops, and one of the reasons for that is that children need to have a vocabulary that makes them able to express feelings and thoughts. For instance, children have to be able to distinguish between mental concepts such as remembering, forgetting, comprehending, guessing and attending (Lovett & Pillow, 1995). In the case of reading, Baker and Brown (1984) concluded that younger readers do not always realize that understanding of texts could be rather effortful. To them the main activity of reading is decoding. Younger readers also seem to lack a sense of reading for meaning. Another difficulty for younger readers is to point out the main ideas in a text (Baker & Brown, 1984). Bouffard et al. (1998) also found that cognitive development affects self-perception such that good performers begin to reflect on their own performance at an earlier age than poor performers. In grade 5 good and poor performers are more equal in their ability to self reflect. The present thesis has acknowledged the fact that metacognition is a skill that gradually develops. Therefore, only high-school students have been recruited as participants in the experiments (Dominowski, 1998).

Powel and Gray (1995) argued that the way children judge themselves in terms of capability is an important indicator of what they eventually will learn and of their motivation to do so. As children grow older they make more accurate predictions of performance, after having gained necessary experience to do so (Powel & Gray, 1995). Schneider (1985) did conclude, however, that even small children can accurately predict their memory of texts given concrete enough tasks. Too abstract tasks and/or lack of task experience is what cause problems for small children' s metacognition. Lovett and Pillow (1996) argued that small children might have problems with metacognitive judgement, but this is partly dependent on what type of judgement they are required to make. The ability to distinguish between different mental processes does not occur until late childhood (Lovett & Pillow, 1995, 1996). Stipek and Gralinski (1996) suggested that fourth graders have developed the cognitive capacities to differentiate between intelligence and performance and to separate ability from effort (Simpson, Licht, Wagner, & Stader, 1996).

Flavell (1979) asked children of different ages to study a set of items and to indicate when they could recall all of them. It was concluded that preschoolers had a limited knowledge about cognitive tasks and behaviors and that they did little monitoring of their own memory. The older children, grades 3 to 5, could point out different variables that affected their memory performance. They knew more about their memory abilities and the fact that remembering varies from time to time and among individuals. Older children also knew that information is lost rapidly from short-term memory if nothing is done to remember and commit the items to long-term memory. Older children indicated that they used some kind of strategy to recall as much as possible. They also indicated more use of mnemonic strategies. For example, they knew the difference between gist and verbatim recall (Garner, 1987).

Brown and Campione (1978) let students from grade 5 to college level read texts. They found that the older readers showed more metacognitive ability in that they could pinpoint the most important parts of a text, and thus where to invest extra effort. Also, Forrester-Pressley and Waller (1984) found that sixth graders knew more about their memory and the need to attend to stimuli to be able to remember them, than third graders. They also found that the older pupils could use their language skills better, discuss memory strategies, and memory skills. This study demonstrated two important things, older learners can use different strategies to improve performance and they can verbalize and monitor their own strategy use. Garner (1987) concluded that preschoolers knew less than older children about different factors that affected their own memory. Even though preschoolers were familiar with expressions such as remember or forget, these were better understood by the older children.

In sum: evidently, problem-solving and other learning situations put hard demands on students if they are to monitor the situations correctly. It seems that a good self-image and an ability to reflect are important prerequisites for metacognitive actions. The present thesis has acknowledged this and only used students who are 15 years or older. At this age students should generally be able to reflect and also know themselves quite well. Also, in most of the present studies the students answered questions regarding their views on reading, memory, and comprehension. These questions were of an open-ended nature and the answers were used to complement and deepen the results of the quantitative data of this thesis.

3. Definition of metacomprehension

Metacomprehension is one area of research underlying the umbrella concept of metacognition. It is for example concerned with our own general knowledge of reading and our ability to evaluate and regulate text processing. One way to tap into students' metacomprehension is to let them rate their current level of comprehension of text and compare these ratings with actual measures of comprehension (Lin & Zabrucky, 1998; Maki, 1998). To make these ratings accurately, students constantly have to take into account familiarity and knowledge about test relevant information, forgetting due to delay, as well as other factors affecting their learning and their understanding of text (Maki, 1998).

Educational psychologists study both evaluation and regulation of cognition (Hacker, 1998; Maki, 1998). Students have to know when they have learned something and when they have not. They also have to know if they need to invest extra effort to better understand text. In order to study regulation of comprehension lexical, syntactic or semantic error detection tasks could be used in which learners are informed or uninformed about inconsistencies of texts. Learners who acknowledge what is wrong with the texts are keeping track of their comprehension. Common measures of error detection are reading times, verbal reports following reading, and having students underline problematic parts of a text (Otero, 1998). Often questionnaires are used to investigate what learners know about their regulation and evaluation of comprehension (Lin & Zabrucky, 1998).

When cognitive psychologists study metacomprehension they focus on evaluation of comprehension monitoring. The main task is to find out whether or not students can accurately evaluate their comprehension of texts (Hacker, 1998; Lin & Zabrucky, 1998). This thesis is based on a cognitive approach towards metacomprehension and has investigated whether the students can accurately calibrate their comprehension of texts or not.

Calibration accuracy is a measure of comprehension monitoring which has been investigated in different ways (Otero, 1998). For example, after having read a text, the readers' ratings of comprehension are correlated with actual comprehension (e.g., number of correct answers to multiple-choice questions). Second, readers' confidence of having answered questions correctly are correlated with actual comprehension. Third, readers' predictions of how well they would do on a comprehension test are correlated with comprehension performance. The higher the relation the better the accuracy (i.e., a positive correlation). The present thesis has used the first measure of comprehension calibration accuracy (Gillström & Rönnberg, 1995; Eriksson, 2000; Eriksson & Rönnberg, 2000).

In their review of calibration studies, Lin and Zabrucky (1998) concluded that correlations generally have been rather low even if some of them have attained significant levels. Earlier studies often used a single item test (e.g., Glenberg & Epstein, 1987), whereas later studies have used multiple test items and thus increased the reliability (e.g., Maki & Serra, 1992; Weaver, 1990).

To sum up, one of the main aims of this thesis has been to investigate how well students can evaluate their comprehension of text. The measure being used is calibration accuracy which indicates whether or not there exists a relationship between students' ratings of their own reading comprehension and actual level of comprehension. The better agreement the better accuracy of ratings.

4. Definition of metamemory

Metamemory is concerned with the relation between students' knowledge about their own memory and memory performance (Carroll & Korunika, 1999; Schneider, 1985). It is valuable that people can identify what has been successfully encoded and what has not, and thus, where to invest extra effort to better remember and learn a material (Begg, Martin &

Needham, 1992). Some metamemory investigations have used single words or word pairs as test material, whereas others, like in this thesis, have used text materials (Cull & Zechmeister, 1994). As stated earlier, children become more aware of themselves and their ability as they grow older (Flavell, 1979; Garner, 1987). Bristow et al. (1998) concluded that at the age of ten students' metamemory is well developed and children know when and where to make an effort to remember. At this point they also have a better understanding about what forgetting something means. Pressley and Schneider (1997) indicated that modern metamemory research have used regression and path analyses which have shown that metamemory measures are strong predictors of performance, and metamemory precedes memory behavior and performance. Thus, a child's knowledge about his or her memory seem to influence strategic behavior which in turn predict memory performance.

According to Carroll and Korukina (1999), there are many possible judgements that could be asked before, during, or after a learning process that in different ways affect recall. Most of these judgements are prospective in nature whereas a few are retrospective. *Ease of learning* (EOL) is an example of prospective judgement in which the students rate the ease with which a certain material has been learned. According to Cull and Zechmeister (1994) these ratings are made rather accurately but Begg et al. (1992) are less certain of the effect that this type of ratings has on recall. They found that people tended to expect high recall of easy-to-process items and that they should forget more difficult ones (Begg, et al., 1992). Ratings are accurate if the factors that lead to ease of learning are similar to those that also causes successful recall. A mismatch between these factors lead to inaccurate EOL ratings (Begg, Duft, LaLonde, Melnick, & Sanvito, 1989; Begg, et al., 1992). Other prospective judgements are those in which students indicate if they have learned a material good enough such that they can remember it for a later test - *judgements of learning* (JOL), (Cull & Zechmeister, 1994). There seem to be a few factors that enhance accuracy of JOLs such as multiple presentation or recollection during study. The everyday experiences of *tip-of-the-tongue* or *feeling-of-knowing* also represent prospective ratings that students often make rather accurately (Carroll, et al., 1997; Schneider, 1985). The prospective ratings of the present thesis is *predictions* by which students indicate how much of the text they will recall (Schneider, 1985). The logic behind this ratings is that if a student has monitored previous performances correctly they should also be able to predict future performances (Pressley & Schneider, 1997). The present thesis also includes one retrospective rating in terms of *postdiction* whereby the students indicate how much of the text they were able to recall (Carroll & Korunika, 1999; Maki, 1998). Both of predictions and postdictions are correlated with text recall and indicate whether or not students show memory monitoring skills (Schneider, 1985). Thus, the better agreement between rated and actual text recall the better prediction and postdiction accuracy.

Koriat (1998) suggests that metamemory constitutes a by-product of memory in that the amount of information that someone can recall usually is rather accurate. In traditional laboratory research, memory has been viewed as a storehouse and students' ability to reproduce information has been the main measure of interest. If someone recalls twenty-seven out of one hundred words this input-bound measure is twenty-seven percent (Koriat & Goldsmith, 1997). In more naturalistic scientific settings, a correspondence metaphor focuses on how many of these recalled words that were on the original list. Koriat and Goldsmith (1997) concluded that what students remember is usually correct indicating high levels of output-bound accuracy. In this situation, consider the fact that if three of the twenty-seven words were not on the list, the output-bound measure would still be ninety percent accurate. Reproduction or accuracy are more or less the same in forced-report situations. In the free-report situation, participants report what they believed to be correct, which usually results in a smaller but rather accurate number of items. Thus, free-report increases participants' monitoring control (Koriat & Goldsmith, 1996a; Koriat & Goldsmith, 1997). In the present thesis, students have recalled the text with their own words with no more help than the title of the text. It should be noted that the students were asked to recall the text as they remembered it, word by word.

Carroll and Korunika (1999) discuss two different views regarding students' metamemory monitoring and the factors that affect people's judgements of their own performance. The first view claims that students make these judgements based on an assessment of the strength of memory traces after acquisition. The other view suggests that besides memory strength, there are other factors involved, such as beliefs about memory per se, and prior experience with the task, that affect judgements. This latter view also discusses extrinsic, intrinsic, and mnemonic cues that are available to a person when judgements are made. Extrinsic cues refer to different conditions of learning, such as level of processing or number of trials. Intrinsic cues could refer to perceived difficulty, and semantic relatedness between words. Mnemonic cue gives information about how well the material was learned, such as familiarity, or outcomes of prior learning. Koriat (1997) claims that judgements of learning are more sensitive towards intrinsic than extrinsic cues. However, Carroll and Korunika (1999) found that both extrinsic (auditory/ visual presentation) and intrinsic cues (ordered/disordered text material) were important for judgements of learning. When their students were given auditory presentation and/or coherent text they recalled more after a delay of two weeks and made higher judgements of learning than they did with visual presentation and/or incoherent text. However, accuracy of JOLs was significant regardless of modality or coherence.

5. Metamemory and metacomprehension combined

The present thesis includes some studies that let students evaluate both their metamemory and their metacomprehension of the same set of texts. According to Pressley, Snyder, Levin, Murray and Ghatala (1987), it is necessary to study both memory and comprehension monitoring. For one thing, most research focuses on whether students can assess their comprehension since understanding text is an important psychological as well as educational goal in learning. However, studying metamemory is as important since it is crucial that a person can rate both the likelihood of remembering and having learnt a material. Also, remembering and comprehending are interrelated, if you have comprehended a text the likelihood that you will be able to recall it is more apparent, even if it is not foolproof (Lovett & Pillow, 1995, 1996). Often comprehension and memory processes are intertwined in everyday life. Typically, good comprehension leads to good memory, yet these are two different mental processes with separate end states (Lovett & Pillow, 1995).

It is difficult to use all kinds of knowledge we have at our disposal effectively, which could result in a mismatch, referred to as "inert knowledge" (Koriat, 1995). Sometimes this term has been used to describe a classroom phenomenon where students may have gained a lot of knowledge in school but they are not prepared for the complexity of real life (Stark, Renkl, Gruber & Mandl, 1998). The psychological criteria for comprehension is to demonstrate a clear representation of the meaning of presented materials, and for memorization it is retention and representation of presented materials. The overt criteria for comprehension could be to get the answer right or to integrate one passage of a story with another. For memorization it is reproduction. To achieve these goals people are required to make the right choices, adjustments, monitoring and evaluation of their work – in other words metacognitive decisions (Lovett & Pillow, 1995). Smaller children seem to have unrealistic ideas of their own performances but as they grow older they make more reliable and accurate ratings of performances (Powel & Gray, 1995). Lovett and Pillow (1996) concluded that not until late childhood can children differentiate between comprehension and memory.

In conclusion, the fourth section described the two main areas of research that constitute this thesis and this, the fifth section, gave the reasons as to why it is important. First, metacomprehension research has as its object to study subjective knowledge and control of the reading comprehension process and reading outcome (Lin & Zabrucky, 1998). In this thesis, the students calibrated how well they had comprehended the text and these ratings were then correlated with actual comprehension measured by answering performance on comprehension questions. Second, metamemory research which investigates subjective knowledge of students' own memory and memory performance (Carroll & Korunika, 1999). As the processing in reading requires both comprehension and memory this thesis has made an attempt to investigate both metacomprehension and metamemory - in parallel, as well as separately (Lovett & Pillow, 1995).

6. Factors being studied

Metacognition is a broad concept, and in this thesis a limited – but crucial - number of factors were investigated. These factors are concerned with *person* in terms of verbal skill, *task* in terms of instructions, time of test, placement of ratings and *text* in terms of different types of text and text characteristics (Lin & Zabrucky, 1998). These factors will be described below but first, it should be noted that throughout the experiments students from upper school levels have participated, that is, grade 9 and high-school students.

The age of the students is important to consider as metacognitive thinking is, as mentioned, a gradually developing skill. Many investigators have used younger children to study how metacognitive behavior develops (Baker & Brown, 1984; Brown & Campione, 1978; Flavell, 1979; Forrester-Pressley & Waller, 1984; Garner, 1987). There were two main reasons why high-school students were used in the present thesis. First, the purpose has been to find out how accurately students can assess their memory and comprehension of texts. From this perspective, high-school students should function well since they have reached the age when they are able to act metacognitively. At this age they have a vocabulary with which they can express cognitive thoughts, they have also gained a variety of cognitive experiences. Second, high-school students are not a well-studied group in this type of research. A great deal of focus has been placed on younger children (cf. Baker & Brown, 1984; Powel & Gray, 1995). Other research has studied adults' metacognitive behavior (cf. Brown & Campione, 1978; Forrester-Pressley & Waller, 1984; Garner, 1987).

6.1 Verbal ability

Students vary in their verbal and reading abilities and this variation has interested metacognitive researchers (Garner, 1987; Lin & Zabrucky, 1998). ”When and how do learners engage in metacognitive actions?”; ”Do all learner engage in metacognitive thinking?”. ”Who is more likely to reflect, plan, evaluate the learning processes?” ”Are these actions more or less accurate depending on the learners intellectual capacities or verbal abilities?”. Lin and Zabrucky (1998) concluded that skilled readers are more likely to use different strategies, and to evaluate and monitor their reading to extract the meaning from texts, than are less skilled readers. Even so, Pressley et al. (1987) argued that this line of research has produced conflicting results when it comes to students' ability to assess their comprehension or memory of texts.

Quite a few studies have come to the conclusion that good performers show better metacognitive ability. Maki and Berry (1984) found that students who scored above the median on a comprehension test more accurately calibrated their delayed performance on this test than students scoring below the median. When they used immediate tests the difference between skill groups was less obvious. Lin and Zabrucky (1998) concluded that high performing students are more likely to engage themselves in conscious processes to improve their learning from texts. Garner (1987) also claimed that good readers were better metacognizers and they begin to behave metacognitively at an earlier age than poor readers. Sinkavich (1995) concluded that high performing students were better self-regulators of their own learning.

Maki and Swett (1987) found that all their students, regardless of skill level, predicted their memory of text accurately. Some of their data indicated that poorer achievers made more accurate ratings. They argued that the test situation was familiar and straightforward for the students in that they read a text and then predicted their recall performance. In this way, all students could monitor the testing situation. Cull and Zechmeister (1994) concluded that methodological factors such as test material, familiarity with the task and performance requirements affect how different verbal skill groups manage to rate their performance.

One reason why different results have been attained is that the definition of what constitutes poor and good readers has varied as well as the age of the learners (Lin & Zabrucky, 1998). In some cases, all participants have been students at higher educational levels. Even if they vary in verbal ability most of them could be regarded as rather good readers. Under these circumstances, comprehension monitoring is not influenced by reading skill (Cull & Zechmeister, 1994; Lin & Zabrucky, 1998). Legree, Pifer and Grafton (1996) found that different cognitive abilities are less correlated among high ability students than in more heterogeneous groups. Cull and Zechmeister (1994) found that two groups of learners who initially varied in their learning ability solved the tasks, and were affected by the tasks in similar manners. Both groups benefited from the presence of test trials during learning and they compensated equally for item difficulty in a self-paced task. Poor learners studied critical items as many times or even more times than good learners, but they recalled them less well.

In the present studies, data have been viewed overall as well as for verbal skill. This way, it could be investigated if and how verbal skill level affected accuracy of ratings. In the first studies, the school teachers were asked to divide students into high and low reading comprehension ability groups. Their division was expected to be validated via objective verbal and memory tests. That is, those regarded to be high on reading comprehension should perform significantly better on tests of antonym-synonyms, analogies, lexical access speed, and reading span. Baddeley, Logie, Nimmo-Smith and Brereton (1985) concluded that the amount of text students can recall is affected by their verbal ability and working memory capacity. Also, according to Jackson and McLelland (1979), lexical access speed is a good predictor of reading comprehension. A second way by which the verbal skill groups were expected to be validated was via performance. Thus, high verbal skilled students should recall more and answer comprehension questions better. From an IQ perspective, these intercorrelations could be due to students' levels of intelligence. As will be shown in Section 8, the teachers accurately divided the students into different verbal skill groups and the groups performed as described above. Therefore, only the verbal test results were used to divide students into different verbal skill groups in Studies II and III.

6.2 Text processing

Winne and Hadwin (1998) claim that metacognitive control requires some knowledge regarding study tactics and strategies. They found that students lacked natural knowledge about effective strategies in that students being trained in study tactics outperformed "normal studying" students (Winne & Hadwin, 1998). Lin and Zabrucky (1998) concluded that different tasks put different effort and cognitive demands on the learner, and an interesting question in this thesis has been to study how different types of text processing affect cognitive as well as metacognitive performance. Maki et al. (1990) for instance, let some of their students read texts with deleted letters whereas others read intact texts. Texts with deleted letters were supposed to increase the level of active reading, that is require more effort, which in turn should increase performance (McDaniel, 1984). The question was if it would improve calibration accuracy as well. In fact, the deleted-letters group excelled in comprehension and made more accurate calibrations. Schommer and Surber (1986) used the same text but varied depth of processing such that some of their students evaluated the clarity of the text (shallow) whereas others prepared to teach someone the main points (deep). The deeper form of processing had a positive effect on students' ability to make test predictions, especially when the students read texts with a high level of difficulty. Yet, Carroll and Korunika (1999) found that coherent texts resulted in better recall and greater magnitude of judgements of learning than non-coherent texts. However, accuracy of judgements of learning did not differ due to the texts being intact or presented in a disordered fashion.

In the present thesis, text processing in terms of reading instructions were manipulated in Experiments two to four (Table 2). In order to vary depth of processing, Experiment 2 let the students assume the roles of learners or teachers while reading school-books and fairy-tales (Schommer & Surber, 1986). The assumption was that reading to teach someone should be a less frequent, different study situation and, thus, more effort requiring (Persson, 1994; Schommer & Surber, 1986). Experiments 3 manipulated level of active text processing and personal involvement (Maki, et al., 1990; McDaniel, 1984) in that the students read texts with (given, selected) or without keywords as extra help. In Experiment 4, the reading instructions emphasized remembering on the one hand and comprehension on the other. Thereby, it could be investigated if students could optimize their reading in both these respects.

6.3 Time of test

In everyday school-life – which this thesis is about – planning and preparing for academic activities is essential which makes it interesting to study time-of-test effects as well as long-term monitoring of memory and comprehension (Hacker, 1998; Lin & Zabrucky, 1998; Sinkavich, 1995). Therefore, some of the students rated how much they would recall and how well they would understand the text in a week's time (Experiments 2 and 4).

Lin and Zabrucky (1998) suggested that there are different factors that affect the accuracy of immediate and delayed ratings of performance. Is it so, that only immediately after exposure ratings can be made accurately, as they are based on immediate impressions? Or, is it the case, that an "illusion of knowing" is more apparent immediately after exposure, as the information from the text is still being held in working memory (Lin & Zabrucky, 1998)? Illusion of knowing is a term suggesting that a person makes ratings based on his or her expertise rather than on actually presented specific text contents (Glenberg & Epstein, 1987). In the Cull and Zechmeister (1994) study, students benefited from short delays in that they spent more time studying difficult items. In all of the present studies, there has been a short delay between ratings and performing, and as mentioned, in two of the experiments there has also been included substantial delays of one week.

Maki and Swett (1987) reported that a week' s delay reduced text recall but both immediate and delayed prediction accuracy was found. Carroll, et al. (1997) used two and six week's of delay and found that their students could not forecast future performance accurately. The students expected the same amount of recall after two and six weeks, but recall was significantly lower after the longer interval. The fact that the students predicted the same recall after two or six weeks was discussed in terms of experimental design. In their study, a between-subjects design was used which could have affected the ability to discriminate between different lengths of delay. They also found that overlearning had a positive effect on text recall compared to semantic relatedness, but that their students thought the other way around (Carroll, et al., 1997). Carroll and Nelson (1993) suggested that studies that includes subjective thresholds, benefit from within-subject designs. It leads to more consistent placements of ratings across different criteria. Similarly, Lovett and Pillow (1996) concluded that a within-subject design make it easier for students to evaluate the effectiveness of reading strategies. In the present thesis both within- and between-subject designs were used. It was assumed that the one-week delay should reduce performance as well as performance predictions.

6.4 Prospective and retrospective ratings

As a measure of memory monitoring, the students of this thesis predicted their text recall, that is, made ratings regarding how well they would be able to recall the text (Table 1). After having recalled the text the students also postdicted how well they actually managed to recall the texts (cf. Carroll & Korunika, 1999). Retrospective ratings, such as postdictions, are usually more reliable than prospective ratings, such as predictions, due to more available information and task-appropriate experience.

As a measure of comprehension monitoring, the students also made prospective and retrospective ratings of their comprehension. The students rated how well they thought they had understood the text and these ratings were correlated with answering performance on comprehension questions (Experiment 2, 3). Retrospectively, ratings of how well they managed to answer the questions were correlated with answering performance on the comprehension questions (Experiment 4).

6.5 Type of text materials

In the study of metacognition, different experimental material have been deployed. Maki and Serra (1992), for instance, concluded that students accurately predicted their recall of lists of words. Extra study time even increased this level of accuracy (cf. Begg, et al., 1992; Lovelace, 1984). In other studies, texts have been used as experimental material (cf. Glenberg & Epstein, 1987; Maki et al., 1990; Maki & Serra, 1992). One of the reasons why text material has been used is to create experimental situations that are more similar to every-day school-life activities, with the aim to increase ecological validity. Furthermore, text material is presumably more demanding than word lists, as it consists of integrated information that has to be understood in a more global sense. The students typically make ratings for longer passages than they usually do with a word list (Maki & Berry, 1984; Maki & Swett, 1987).

The four experiments that laid the groundwork for this thesis were based on different types of text material (Table 2). In Experiment 1 short, easy-to-read stories were written by the experimenter. These texts varied in their consistency and distinctiveness (Maki & Swett, 1987). Experiment 2 used fairy-tales and school-book texts (Persson, 1994). These texts were considerably longer compared to those in Experiment 1. The two texts used also differed, such that one of them emphasized reading for pleasure and the other reading for learning (Persson, 1994). Expository texts were used in Experiments 3 and 4. These texts were taken from a reading comprehension test and in these experiments, focus was placed on instructional effects (Glenberg & Epstein, 1987). One of the differences between narratives and expository texts is that the former is written to entertain and encourage readers to attend to global ideas of a theme, whereas the latter is written to communicate information and encourage readers to attend to details of a text (Carroll & Korunika, 1999). It has been shown that performance predictions should be based on thematic questions for narratives and detailed questions for expository texts to be accurate. However, if equated for difficulty level, it is texts of intermediate level of difficulty and not type of text that contribute to accuracy of ratings (Carroll & Korunika, 1999).

7. Design and purpose

Ninth graders and high-school students participated in this thesis. The students calibrated their comprehension of text and predicted their text recall. Positive and significant correlation between these ratings and actual text recall and answering performance were taken as an indication of accuracy, that is, correct memory and comprehension monitoring. The students also made retrospective ratings of comprehension and memory – postcalibration (Experiment 4) and postdiction (Experiment 3, 4) accuracy of texts. Data have been analyzed overall but also for different verbal skill groups. Personal involvement, activity level, and depth of processing have been manipulated via the use of different types of instructions and texts. The students have made both on-line and long-term ratings of performances.

Table 1 describes the chain of events in the experiments of this thesis. The participants have read a text and *after* that, predicted their memory of texts and/or calibrated their comprehension thereof. After a short while they recalled what they remembered of the text and also answered multiple-choice questions regarding their comprehension of the text. In the last two experiments the students were also asked to postdict and postcalibrate their performances. In all experiments, the participants have been informed that they are going to make different types of ratings and also that they should recall the text and answer questions.

Table 1. *Show the Overall Chain of Events in the Experiments.*

Metamemory

Immediate testing	*Delayed testing*
prediction - recall - postdiction - predict delayed recall	recall - postdiction

Metacomprehension

Immediate testing	*Delayed testing*
calibrate – answer – postcalibration calibrate delayed performance	answer-postcalibration

7.1 Overall design

The main purpose of this thesis has been to investigate how accurately high-school students can evaluate their memory and comprehension of text. The assumptions being made in this thesis center around *metacognitive knowledge* in terms of person, task, strategy and text related variables (Flavell, 1979; Garner, 1987; Lin & Zabrucky, 1998). If students can evaluate their performance there should be significant relation between rated and actual performance – calibration and prediction accuracy (Hacker, 1998; Lin & Zabrucky, 1998; Maki, 1998). The overall design of this thesis is summarized in Table 2.

Table 2. *Shows the Overall Description of the Experimental Conditions.*

Text	**Instruction**	**Ratings**	**Time of test**	**Design**
		Experiment 1		
story	general reading	prediction	immediate	between-subject
		Experiment 2		
school-book fairy-tale	learn teach	prediction	immediate delayed	Between-subject
		Experiment 3		
expository	reading given selected	prediction postdiction calibration	Immediate	within-subject
		Experiment 4		
expository	understand remember	prediction postdiction calibration post-calibration	immediate delayed	within-subject

As was shown in section 6.1, verbal skill is a complicated matter in metacognitive research but it was expected that high verbal skilled students should recall more of the texts and answer the comprehension questions better than low verbal skilled students (Baddeley et al., 1985; Jackson & McLelland, 1979). In detail however, it could be expected that verbal skill could result in two possible outcomes:

- There are *no differences* between the verbal skill groups. The students could all be regarded as experienced readers as they were 15 years or older and had been going to school for 9 or more years. This could result in similar levels of accuracy ratings (Cull & Zechmeister, 1994; Lin & Zabrucky, 1998).
- High performing students should make *more accurate ratings* than low performing students, as they are better metacognizers

and have more metacognitive knowledge (Garner, 1987; Sinkavich, 1995).

The task aspect of this thesis concerns instructions, time of test, and placement of ratings. Thus the following assumptions were made:

- Instructions emphasizing learning should be regarded as effortful and thus, result in better cognitive performance and accuracy of ratings (McDaniel, 1984; Persson, 1994).
- The more personal involvement and the more effort delivered in solving the task, was expected to have a positive effect on cognitive performance and accuracy of ratings. Having to use keywords as extra support should be more demanding compared to reading only. Especially when the students selected their own keywords (Maki, et al., 1990, McDaniel, 1984; Schneider & Laurion, 1993).
- The students make ratings of immediate performance more accurately than ratings of delayed performance (Carroll & Nelson, 1993; Carroll, et al., 1997).
- Retrospective ratings are made more accurately than prospective ratings, as the former is based on task-appropriate experience. Thus, postdictions and postcalibrations are more accurate than calibration and prediction accuracy (Maki, 1998).

The text aspect of this thesis has resulted in some general assumptions:

- A Von Restorff effect is expected such that inconsistent texts result in better cognitive performance and accuracy of ratings (Maki & Swett, 1987).
- The reading of school-book texts or expository texts is regarded as more effortful than that of fairy-tales or narratives, resulting in better cognitive performance and accuracy of ratings of the former types of texts (Maki, et al., 1990).

38

7.2 Open-ended questions

Due to the cognitive approach of this thesis, the key emphasis is on monitoring, that is, to what extent do students know their current - immediate or delayed - state of remembrance and understanding of a text. At the end of Experiment 3 and 4 the students answered some open-ended questions in which they evaluated the usefulness of the instructions. These evaluations made it possible to collect some qualitative data in addition to the main collection of quantitative data. Otero (1998) suggested that a combination of quantitative and qualitative data gives a better overview of the results and also a possibility to deepen the interpretations (Otero, 1998). Examples of these questions are *"What do you think is typical of a good reader"?*; *"Which instruction did best facilitate your text recall?"*; *"What do you usually do when you have to remember something"?*.

8. Description of Study I-III

This thesis comprises of three studies that consist of four experiments (two experiments in Study I). Section 8 describes these studies, their results and the discussions following the results. This section ends with a summary of both consistent and inconsistent data patterns.

8.1 Study I

The first study consisted of two experiments that investigated prediction accuracy of text recall. In Experiment 1, four easy-to-comprehend short stories were used, written by the author of this thesis. These texts varied in distinctiveness and consistency (Maki & Swett, 1987). Johnson (1970) suggested that texts that are coherent and consistent relative to a schema are better recalled than inconsistent text materials. The von Restorff effect, on the other hand, suggests that irrelevant information improves text recall. Maki and Swett (1987) tested these two perspectives on prediction accuracy in that their narrative texts varied in consistency. In some aspects the results indicated that the Maki and Swett's students had a von Restorff basis for their predictions. Text recall was higher for inconsistent texts but prediction accuracy did not differ due to consistency, correlation coefficients were significant in both cases. Experiment 1 in the present thesis included distinct items of texts as an additional von Restorff effect. In this experiment the term "distinct" was used to describe striking and/or dramatic textual contents (Gillström & Rönnberg, 1994).

Table 3 shows the four experimental conditions under which 80 ninth-graders were tested (20 students in each cell). The students varied in their reading ability such that half of them were regarded as high and the remaining students as poor readers (10 good and poor readers in each cell). This division of students was made on the basis of teacher ratings. In later studies (Studies II and III), the term low, medium and high verbal skilled students were used whereas Study I used the term reading ability.

Table 3. *Shows the Conditions in Experiment 1.*

condition 1	*condition 2*	*condition 3*	*condition 4*
distinct	non distinct	distinct	non distinct
consistent	consistent	inconsistent	inconsistent

Based on Maki and Swett (1987) it was assumed that an item or idea in a text that is perceptually and cognitively different from the schema should yield better cognitive performance as well as metamemory of texts (cf. Grasser, Woll, Kowalski & Smith, 1980). Thus, inconsistent texts should result in better text recall and better awareness of what one will remember from a text than consistent text. It was assumed that a text that contains both distinct and inconsistent contents should increase the von Restorff effect even further. From this the following combinations of text characteristics and assumptions were made:

- Text 1 was **consistent-distinctive** about a boy who was going to visit his brother. When the conductor was coming towards him he could not find his ticket which made him feel really uneasy *(distinct episode of the text)*. This combination of distinct elements and consistency, should result intermediate text recall and prediction accuracy of texts together with text 4.

- Text 2 was **consistent-non distinct** about two boys who preferred to see documentary films and they went to see a film of that kind. This combination should be the most "boring" text and result in the lowest levels of text recall and prediction accuracy of texts
- Text 3 was **inconsistent-distinct** about a boy who was going to visit his brother. He stood and waited for the train and suddenly sat in a car *(inconsistent episode)*. When the conductor was coming towards him he could not find his ticket which made him really uneasy *(distinct episode)*. This combination should result in the best possible text recall and prediction accuracy as it was both inconsistent and distinctive.
- Text 4 was **inconsistent-non distinct** about two boys who preferred to see documentary films but went to see a love story *(inconsistent)*. The inconsistent elements of this text should together with text 1 result intermediate text recall and prediction accuracy of texts together with text 1

To validate the text manipulations the students rated level of distinctiveness and consistency of the text. Accurate ratings would indicate that the students identified and correctly used the available information of the text characteristics.

Our results showed that the students could identify the text variations in that their ratings of distinctiveness and consistency were in agreement with the actual level of text variation. Thus, distinctive texts were regarded as more distinctive than the non-distinctive and consistent texts as more consistent than inconsistent texts. However this text variation per se, had no effect on ratings, recall or prediction accuracy. On the whole, no subjective prediction accuracy was found which was discussed in terms of ease of processing (Begg, et al., 1989). The texts could have been too easy to read which made the students spend too little time studying the texts for remembering purposes. To prove this point, additional data analyses were carried out suggesting a clear pattern of overestimation. That is, the students believed that they would recall more than they actually did.

Furthermore, verbal skill differences were found in text recall performance (Paris & Meyers, 1981) but again, not in accuracy of ratings. high verbal skilled students recalled more text than low verbal skilled students but none of these groups made more accurate ratings than the other.

However, the first experiment yielded *objective prediction accuracy* which is a result that has been replicated throughout this present series of studies (Eriksson, 2000; Eriksson & Rönnberg, 2000; Gillström & Rönnberg, 1994, 1995). This type of accuracy reveals that objective measures, such as verbal test results (antonym/synonym, analogy) or a working memory test (reading span), correlate positively with text recall. Thus, students who score well on these tests also recall more of the texts than low scoring students.

The result of the first experiment formed and motivated the remaining set of studies in this thesis. Thus, to improve subjective accuracy of ratings it was regarded as important to increase the effort demands. The students had to study texts more closely and for a longer period of time to assure prediction accuracy of text recall (O´Brien & Meyers, 1985; McDaniel, Einstein, Dunay & Cobbs, 1989). Hence, the second experiment tested the idea that prediction accuracy requires active text processing (O´Brien & Meyers, 1985; McDaniel, et al., 1989). It was assumed that performance as well as knowledge about performance should benefit from more conscious and deliberate processing of texts. McDaniel (1984) found that students who filled in deleted letters while reading a text recalled more of the text than those who read intact texts. Maki et al. (1990) showed that filling in letters improved calibration accuracy as well. Presumably due to increased effort demands.

In Experiment 2, the text processing demands were increased by asking the students to underline words or sentences that they found important based on one of four reading situations (see Table 4). In this way they formed their own key for recall. The texts that were used was also longer than the ones in Experiment 1 and, in addition, the students participated at a delayed testing, one week after the text reading session.

Experiment 2 also investigated familiarity with a study situation. Persson (1994) found that students were more familiar with school-book texts and they also found this type of text material more demanding. One argument was that school-book texts are more associated with learning, compared with other types of texts, such as fairy-tales. The students in Experiment 2 were instructed to assume either the roles of learners or teachers and they read either fairy-tales or school-book texts (Table 4). A total of 129 students from grade 9 of the Swedish compulsory school participated. Teacher ratings were used to divide the participants into three verbal skill levels and 44 of them were regarded as high , 49 normal and 36 as low verbal skilled students. Verbal skill level analyses were carried out pooled over experimental conditions.

Table 4. *Shows the conditions in Experiment 2.*

condition 1	condition 2	condition 3	condition 4
learn	teach	learn	teach
school-book	school-book	fairy-tale	fairy-tale
(LS, 29 students)	(TS, 31 students)	(LF, 34 students)	(TF,35 students)

The experimental conditions were expected to vary in level of familiarity and effort demands. To validate these expectancies the students rated how often they read this type of text and how effort demanding they had found the reading situation to be. Again, correct ratings would measure that the students had task relevant knowledge (Flavell, 1979). It was assumed that the most familiar and effort requiring condition should receive the best possible text recall and metamemory of text. From this context one should expect that:

- **LS** should result in best text recall and prediction accuracy of text as this should be the most familiar situation and most effort requiring due to its clear relation to learning.
- **TF** should result in lowest levels of text recall and prediction accuracy of text as this situation should be the least familiar and

least effort requiring reading situation. It had no clear relation to learning.

- **TS** and **LF** should result in medium levels of text recall and prediction accuracy of text as they contained one familiar and effort demanding part each in terms of either learning or school-book texts.

The results demonstrated that three of the experimental conditions received immediate prediction accuracy, only in the TF-condition no significant prediction accuracy was obtained. The students were asked to rate both familiarity and effort requirements as regards type of text/instruction. The ratings indicated that they were most familiar with and also found LS-condition the most effort requiring. This instruction yielded reliable prediction accuracy even after a week's delay. That is, at the immediate testing the students could accurately predict what they would remember from the texts in a week's time.

Viewed from a verbal skill perspective, high achievers students found the text less effort requiring, than normal and low achievers (in this experiment the terms high, normal and low achievers were used). They excelled in performance but were unable to predict their immediate as well as delayed text recall. Normal and low achievers showed either immediate or delayed prediction accuracy.

In sum: viewed from a verbal skills perspective, data indicated that high achievers excelled in performance - but surprisingly - they made less accurate predictions of performance compared with lower achievers - presumably because processing of text and task was less effort requiring for them. Our first study indicated that active text processing is one prerequisite in this type of research. If the text is too easy to read students tend to spend too little time reading the text to remember it and to be able to make accurate ratings (McDaniel, et al., 1989; O´Brien & Meyers, 1985). Reading in order to learn the contents of a school-book text was regarded as the most familiar and at the same time most effort requiring condition which yielded both immediate and delayed prediction accuracy of texts.

8.2 Study II

In Study II, high-school students both calibrated their comprehension and predicted their recall of the same set of texts. Pillow and Lovett (1995) suggested that remembering and understanding the contents of texts are two separate but closely intertwined mental processes. It is easier to remember something that is well understood and vice versa even if it is not absolutely necessary. Study II was also based on the results of Study I. Performance prediction studies seem to require allocation of attention which our data suggested could be attained by increasing the effort demands.

In this study, level of effort was varied in terms of reading strategies. According to Wade and Trathen (1989) it is not easy to find one study technique that would optimize recall or comprehension of texts. For instance, Maki and Serra (1992) found that practice before taking a test had no effect on comprehension or accuracy of ratings. However, increased personal involvement and more active and effortful conditions have generally been found to improve performance as well as accuracy of ratings. Examples of these are conditions that require high involvement (Schneider & Laurion, 1993), cue-review (Begg, et al., 1992), or deletion of letters (Maki et al., 1999).

In study II, a within-subjects design let students use three different instructions (Carroll & Nelson, 1993). All of them emphasized reading to understand, but instructions two and three also contained key-words. These key-words were either *given* to the students or *selected* by the students themselves, and the students could use these key-words during reading, rating, recalling the text and answering questions. An increase of effort demands was expected from instruction one to instruction three which in turn should lead to improved performance and more accurate ratings in the third condition (Maki, et al., 1990; McDaniel, 1984). A total of 111 high-school participated. They were divided into different verbal skill groups on basis of their verbal test results. So as to let extreme skills be analyzed, the top 20 students were regarded as high, 72 as medium and 19 as low verbal skilled students (cf. Sinkavich, 1995). The following assumptions were made:

- the least active and effort demanding task should be requested in the READING-condition, that is reading a text without using any key-words. This should result in the lowest cognitive performance and least accurate predictions and calibrations.
- medium level of activity and effort should be requested in the GIVEN-condition, that is, to read a text having to use given key-words. This instruction should result in medium cognitive performance and accurate performance predictions.
- highest level of activity and effort should be required in the SELECTED-condition, that is, to read a text having to select and use key-words. This instruction should result in highest cognitive performance and accurate performance predictions.

It should be noted that regardless of instructions the students were given the same amount of reading time (i.e., 4 minutes). The students rated how effort requiring they had found each of the experimental conditions to be.

Table 5. *experimental conditions in Experiment 3.*

condition 1	condition 2	condition 3
read to understand	read to understand	read to understand
READING	+ use GIVEN	+ use SELECTED
	key-words	key-words

As expected the subjects' ratings of effort differed such that instructions with given and selected key-words were rated as more effort requiring than the one without key-words. This difference in effort ratings did not have the expected effect on performance and accuracy though. Regardless of instructions, the students answered the questions as correctly, recalled the same amount of text, and accuracy of ratings was the same. The reason for this unexpected result was given by the students themselves in that they indicated that they preferred different ways to study texts. The students were asked to mark which instruction they would use again if they were to read another text for remembering and comprehension purposes. Their marks showed that some students preferred to study texts without any key-words. To them, key-words are obstructing their own learning. Others preferred to be given some key-words which let them concentrate on reading but at the same time they gained some extra help in terms of key-words. The third group preferred to SELECTED their own key-words, they argued such that what improves their learning might not be regarded as important by others.

When data were analyzed from this study preference perspective, the best possible text recall but lower levels of prediction accuracy were received for the preferred instructions. The group who claimed that they recalled the most with READING also recalled more with this instruction than the same group did with SELECTED or GIVEN. The same result was received for those who preferred either SELECTED or GIVEN. It is important to note that this result was not obtained for comprehension of text. The students did demonstrate different preferences for comprehension as well. A clear majority indicated that the READING instruction best facilitated their understanding of text but neither their answering performance nor their calibration accuracy could confirm this result.

To sum up, some data in Study II suggested that reading to comprehend and reading to remember represent two qualitatively different mental activities that are affected in different ways by the same set of variables. If we look at comprehension first, students made significantly lower calibration ratings with SELECTED but actual answering performance did not differ due to instruction. Furthermore, the students could not evaluate which reading strategy worked best for their comprehension. In Study II, it was discussed that comprehension tasks should not be restricted by time and that comprehension ratings might be influenced by familiarity and social factors. For instance, low verbal skilled students are expected, and expect themselves, to do poorly on reading tasks (Persson, 1994). The pattern for recall was different, instructions did not affect either ratings or performance, but more importantly, students could evaluate the reading instructions being used. Verbal skill differences in accuracy of ratings were more clearly found for memory of texts than comprehension of texts, especially when postdiction accuracy were considered. Another important result in Study II, was the fact that the students rated their general memory and reading comprehension abilities with some degree of accuracy. The high verbal skilled students made higher ratings compared to the lower verbal skilled students. This could be taken as evidence of social influences – low verbal skill know that they are poor performers and high verbal skill that they are good performers (Guthrie & Kirsch, 1987; Rueda & Mehan, 1986).

8.3 Study III

Study III investigated whether or not focusing on one mental process at a time would have a selective and positive effect on ratings of comprehension and recall and the related metacomprehension and metamemory (Lovett & Pillow, 1995, 1996). The students read two texts, one with the instruction to remember as much as possible, the other to comprehend it as well as they could. If the students followed instructions, best possible comprehension should be received with the 'understand' instruction and best possible recall with the 'remember' instruction (Table 6). The students were tested immediately and after a delay of one week and they made prospective as well as retrospective ratings of their performance. Second, Study III investigated free versus forced reading time. Study II used the same type of texts and gave each student four minutes of reading time. For control purposes, it was regarded important to study students' behavior in a free reading situation. The hypothesis was that reading time should have a minor effect on the test situation (Mazzoni & Cornoldi, 1993). A within-subject design was used, a total of 88 students participated in the immediate testing and 68 of those individuals participated a week later. Twenty percent of these were regarded as high and low verbal skilled students, respectively.

Table 6. *Shows the Design of Experiment 4.*

type of reading	condition 1	condition 2
free reading time	read to REMEMBER immediate/delayed	read to UNDERSTAND immediate/delayed
forced (four minutes)	read to REMEMBER immediate/delayed	read to UNDERSTAND immediate/delayed

In general, the results showed that focus on one processing at a time had a positive effect for recall but not for comprehension. Calibration and prediction accuracy did not differ due to instructions, overall significant correlations were attained for both experimental conditions. Further and as expected, forced or free reading time had no significant effect on either performance, ratings, or accuracy of ratings. Verbal skill effects were found, showing that high verbal skilled students excelled in performance but again showed less accuracy of ratings. For the first time there was one exception however, high verbal skilled students demonstrated immediate postdiction accuracy when they read to remember. One reason could be that high verbal skilled students were those who best could follow the instruction as they recalled 50% of the text with REMEMBER and only 38% with COMPREHEND. That is, they acted according to instruction and also evaluated their performance with accuracy.

8.4 Summary of data patterns

A summary of data patterns suggests five important results. First, at an overall level, students have an ability to predict their text recall performance with accuracy. Second, the thesis has not been able to present a similarly clear picture for students' ability to calibrate comprehension with accuracy. Calibration accuracy of comprehension seems to depend on a combination of several subtle methodological factors. Third, from a skills perspective, it has been shown that students with lower verbal ability make more accurate ratings of performance, especially for postdictions. Fourth, students' ratings of evaluation are more accurate if they are made after rather than prior to performance. Hence, thinking backwards is more efficient than thinking forwards. If there is a long delay of one week between reading and performing, accuracy of performance predictions are reduced and so is actual text recall but not answering performance. In addition, postdiction accuracy of text recall was significant after a delay of one week. Fifth, the thesis has varied type of reading strategies with text and it has been shown that experimental situations which have been familiar and/or effort demanding have yielded better accuracy of ratings than conditions considered to be less familiar and/or effort demanding. The use of reading strategies also resulted in a study preference perspective on metacognition, determining recall performance and predictions.

9. Discussion

The study of metacognition aims to find out what a person knows about his or her cognitive abilities, and how he or she manages to solve cognitive tasks with best possible outcomes. Thus, metacognition is concerned with how to plan, regulate, monitor and evaluate cognitive performance. Metacomprehension narrows this down to subjective knowledge about, for example, reading comprehension and the ability to evaluate and regulate text processing (Hacker, 1998; Lin & Zabrucky, 1998). In turn, metamemory is concerned with people's knowledge about their own memory and memory performance (Schneider, 1985).

To recapitulate, this thesis has had a cognitive focus which resulted in three studies concerned with students' monitoring of their own comprehension and memory of texts. After having read the texts, the students rated *"how well they thought they had understood the text"*, and also *"how much of the text they thought they would be able to recall"* (Schneider, 1985). If students can evaluate their comprehension and memory of text it should show via calibration and prediction accuracy in the form of significant and positive correlations between rated and actual performance. This thesis is based on variables similar to those in

Flavell's model of cognitive monitoring, that is, metacognitive knowledge in terms of *person*, *task*, *strategy* and *text* (Flavell, 1979; Garner, 1987; Lin & Zabrucky, 1998).

9.1 Metamemory and metacomprehension - results

Three of four experiments in the present thesis yielded prediction accuracy of text recall (Eriksson, 2000; Eriksson & Rönnberg, 2000; Gillström & Rönnberg, 1995). The pattern obtained for calibration accuracy was not as straightforward. Gillström and Rönnberg (1995) showed that students calibrated their comprehension with some accuracy, whereas the more recent studies did not (Eriksson, 2000, Eriksson & Rönnberg, 2000). The results attained for prediction and calibration accuracy are consonant with other research (Maki & Swett, 1987; Schneider & Laurion, 1993). When it comes to metamemory, Pressley and Schneider (1997) summarized over a hundred metamemory studies and found an average correlation coefficient of $r = .41$ between predictions and actual recall. In our studies prediction accuracy varied between $r = .30$ and .55, which seems to be reasonable compared with the Pressley and Schneider (1997) data. Postdiction accuracy of text recall has reached r's between .40 and .70, which also seems reasonable as retrospective ratings usually are more accurate than prospective ratings (Maki, 1998). In an additional study, not reported in this thesis, it was shown that postdiction accuracy occurred even after a delay of one month (Eriksson & Rönnberg, 2000). That is, one month after having read the text the students recalled what they remembered and then made accurate ratings of how well they managed to recall the texts they read one month earlier. It seems that the students have a well kept conception of the text to relate to even after a delay as long as one month.

One of the reasons why both prediction and postdiction accuracy of text recall were obtained seems to be associated with effort. The conditions in the first experiment (Study I) were too easy and as a result no prediction accuracy was found. In the second experiment (Study I), attempts were made to increase the demands in terms of (1) extended text materials, the students either read a fairy-tale or a school-book text, (2) the students having to create their own key for recall, and (3) the use of different instructions (i.e., to learn or to teach). Being instructed to learn a school-book text (LS) yielded both immediate and delayed prediction accuracy. This condition was regarded as the most effort requiring and familiar by the students. Teaching someone a fairy-tale (TF) did not result in any accuracy of ratings and this condition was regarded as significantly less effort requiring than LS. Teaching someone a school-book text and learning a fairy-tale demonstrated immediate but not delayed prediction accuracy.

In study II, the mean effort ratings ranged between forty and fifty percent, indicating a substantial level of task demands. Hence, prediction accuracy occurred. The study preference analyses in Study II also showed that when students studied texts in their personal, best way, they recalled the most, but levels of prediction accuracy were reduced. Regardless of verbal skill, students seem to become skilled readers in the sense that they develop personal strategies and are aware of them. Torrence, Thomas and Robinson (1993) also found that their students were attracted to different types of instructions. Taken together, a general conclusion is that as long as readers are dealing with concrete and demanding tasks, memory awareness is enhanced and more explicitly demanded. Proficiency, due to long practice or skill, requires little effort and awareness, yet resulting in best possible text recall. In this vein, Logan (1988) suggested that novices' responses are at first based on conscious, sequential processes governed by rules, whereas experts rely on memory-based representations. Also, Ackerman (1995) argued that ability-performance relations can be segmented into three broad stages of practice – cognitive, associative, and autonomous. The first stage is mainly associated with novel task performance on verbal and numeric tests. The second stage is associated with perceptual speed abilities, and the third stage is associated with psychomotor abilities.

When it comes to metacomprehension, Lin and Zabrucky (1998) concluded that calibration accuracy measures are sensitive towards methodological matters. In Studies II and III, the text materials were taken from the same reading comprehension tests but the number of participants differed such that there were twice as many students in the former study. This fact could have led to a wider range of calibrations among the high, medium and low verbal skill students in the former (Gillström & Rönnberg, 1995) but not in the latter studies (Eriksson, 2000; Eriksson & Rönnberg, 2000). Tendencies were in the expected direction. That is, the high verbal skill students made higher ratings and better answering performance than the lower verbal skill students. Another methodological aspect of the present studies regards the number of questions being used to measure comprehension. Lin and Zabrucky (1998) suggested that comprehension is a continuous variable and should be assessed by multiple questions. Weaver (1990) found that correlations between rated and actual performance gradually increased when 1,2 or 4 questions were used to test comprehension. The expository texts being used in the present studies were rather short and contained two questions. The proportion of answering performance was correlated with ratings made on a 100 % scale which could have biased calibration accuracy estimates. It should be noted though, that the two questions separated the verbal skill groups. That is, high verbal skill students were clearly much better at putting a title to the text, summarizing the main points into a single sentence, and so forth, than the lower verbal skill students. If longer texts and/or more questions had been used, it would seriously have prevented the combined study of metamemory and metacomprehension, especially in within-subject designs.

There are also theoretical aspects to consider. First, calibration accuracy measures the relation between ratings of comprehension and answering performance and it could be that a person feels he or she has understood the text but his or her personal understanding does not correspond with the aspects of comprehension that the questions address. Thus, there could be confusion between the question-makers' expectations and the learners' intent (Eriksson & Rönnberg, 2000; Wenestam, 1993). Benjamin, Bjork, and Schwartz (1998) found that students sometimes fail on tests even when they feel ready, which could be due to their criteria of learning not matching the actual task demands. Hallam and Francis (1998) let experienced readers read texts from different knowledge domains. They found that students' understanding and appreciation of texts varied, even when they read a text from their own domain. It was concluded that matters such as interest in text and prior knowledge affected comprehension and it should not be expected that everyone agrees on the criteria of comprehension (Hallam & Francis, 1998). Thus, reading comprehension could be much of a personal matter.

Second, Spiro and Meyers (1984) suggested, that*"knowing that you know"* is a more demanding task than the students sometimes realize. As a reader, you have to be able to sort out relevant from irrelevant information, pinpoint main idea, what part needs extra study, and so forth, which could take more time than what the experimental situation provides. Conway, Gardiner, Perfect, Anderson, and Cohen (1997) followed a group of students during a course and found that the students shifted from remembering to comprehending the material. The Eriksson and Rönnberg (2000) data support this result, in that a month's delay reduced text recall to a minimum, whereas comprehension was not affected as much. Across instructions, as many as 18% answered more comprehension questions and 40% the same number of questions correctly after a month. Many students also underestimated their answering performance after a month. None recalled more after a month, reduction was inevitable and clear. Less than 20% of the texts were remembered and more over-estimations were made (Eriksson & Rönnberg, 2000).

Third, it is also necessary to address the issue of social aspects on data. In Gillström and Rönnberg (1995), the calibration ratings differed such that the students expected poorer comprehension with SELECTED compared with the other two instructions (i.e., READING and GIVEN). However, this expected reduction was not confirmed, the students understood the texts equally well regardless of instruction. This could suggest that calibrations are based more on verbal knowledge and familiarity than on actual comprehension. Schneider and Laurion (1993) found that their students relied on familiarity with the topic rather than actual content (cf. Glenberg & Epstein, 1987). How students view themselves could also influence calibration ratings. Guthrie and Kirsch (1987) suggested that good and poor readers are treated differently in school and this could affect their ratings of comprehension (Karabenick, 1996; Persson, 1994). In fact, school research has shown that children who view themselves as low-ability students seem to expect failure after failure, compared to medium- or high-ability students who explain failures in terms of insufficient effort or task difficulty (Helmke & Aken, 1995; Simpson et al., 1996).

In this thesis, one attempt was made to improve students' ability to evaluate their comprehension of text in terms of postcalibrations of their answering performance. That is, the students rated how many of the questions they managed to answer correctly. If calibrations are more person related postcalibrations typically let the students focus on the task (Benjamin, et al., 1998; Hallam & Francis, 1998; Wenestam, 1993). Postcalibrations are similar to the postdictions which supposedly are more reliable and accurate as they are based on task-appropriate experience (Lin & Zabrucky, 1998; Maki, 1998). An alternative way to improve calibration accuracy could be to go from forced- to free-report situations. Tests of comprehension are usually forced in that the same questions are used to test comprehension for each participant (Hallam & Francis, 1998). When students are asked to recall text, they report what they remember in their own words. It usually results in a smaller but accurate number of items (Koriat, 1995; Koriat & Goldsmith, 1997). In the future, an experimental design that would use free-report of comprehension could be one way to increase calibration accuracy.

To conclude, this thesis suggests that the students can rely on their metamemory when it comes to memory monitoring, measured via prospective and retrospective ratings of text recall. The present thesis suggests different reasons as to why the students cannot monitor their comprehension equally well (these results will be thoroughly discussed in 9.3).

9.2 Additional analyses and validity aspects

This thesis has reported a large body of data and at this point, four different aspects of validity will be reported, beginning with potential effects of gender. Other research has not provided any systematic differences when it comes to male and female students and their cognitive and metacognitive performance. Lovett and Pillow (1995) did not find any gender differences in children's ability to distinguish between comprehension and memory. Otero and Campione (1992) found no differences in male and female students' metacomprehension monitoring ability. First, in the present studies, no gender differences worthwhile reporting have been found. To present pertinent evidence in this respect, the male students in Experiment 3 recalled 42% whereas the female students recalled 43%. Both groups made reliable ratings of text recall, *r*´s centered around .40 ($p< .01$). In Experiment 4, no differences in text recall or answering performance were found, but the female participants made reliable predictions for one of the instructions whereas the male participants showed prediction accuracy for both instructions. Also for comprehension, no gender differences were found. Their answering performance was similar and both groups made relatively unreliable calibrations but reliable immediate postcalibrations of comprehension.

Second, the data in Experiment 3 were selected at a time when Swedish students spent between two and four years in high-school (today everyone studies for three years). Those who studied for two years were usually trained for a specific job, such as hair-dresser or office-clerk. These vocational studies usually had a practical focus whereas the three and four year programs emphasized more theoretical knowledge. Thus, one way to address the issue of validation is to show how these different study-groups performed in the experiments (Table 5). As can be seen, significant one-factor ANOVAs were found for average grades, text recall and answering performance with READING, indicating that the three-year students had higher grades and performed better than the vocational students. The same result was attained with GIVEN and SELECTED. Top average grade at the time was 5.0.

Table 7. *Shows the Two- and Three-year Students' Average Grades, their Recall Performance and Answering Performance.*

Group	**n**	**study length**	**average grades**	**text recall**	**answering performance**
Consumption	8	(2 year)	2.69	31%	35%
Social	15	(2 year)	3.20	27%	50%
Administration	13	(2 year)	2.91	35%	56%
Economy	18	(3 year)	3.14	51%	67%
Civics	25	(3year)	3.05	54%	74%
Humaniora	16	(3 year)	3.35	43%	81%
Science	16	(3 year)	3.70	46%	75%
Anovas			$F(6, 83) =$	$F(6, 104) =$	$F(6, 104)$
* $p < .05$			= 4.94*	= 6.95*	= 3.48*

Skolverket (1996) reported about the IALS studies which investigated literacy among adults. It was shown that the lower the educational level, the poorer the literacy skills (i.e., reading, writing and calculating ability). Thus, it seems reasonable to expect that two-year students should have poorer reading and verbal performance compared to three-year students. Table 5 confirms this reasoning, which becomes even clearer with additional ANOVAs on the antonym/synonym tests, $F(6, 110) = 16.89, p < .01$; general ratings of fluency of reading, $F(6, 110) = 4.88, p < .01$; reading comprehension, $F(6, 110) = 4.15, p < .01$; and memory abilities, $F(6, 110) = 2.18, p = .05$. Hence, two-year students' word knowledge is weaker and they also regard their general reading and memory abilities as poorer than three-year students. These measures can also be part of a reasoning with regard to potential effects of general intellectual abilities (IQ). Anderson and Freebody (1985) claimed that the relationship between vocabulary and general intelligence is very strong. The present studies have shown that the word knowledge test being used consistently correlates with text recall and memory skill. In addition, the verbal skill groups have differed in ratings as well as in performance.

Third, in Experiments 3 and 4 the students rated their general reading and memory abilities (Eriksson, 2000; Gillström & Rönnberg, 1995). Thus, the students rated their ability to read fluently, their reading comprehension ability and ability to remember text. These subjective ratings correlated significantly with objective measures such as verbal test results, recall performance and answering performance. Those who made high ability ratings also performed well and vice versa. These data have been taken as evidence that the students can make reliable ratings of their own performance (Cull & Zechmeister, 1993; Wade & Trathen, 1989).

Fourth, in the first two studies teachers were asked to distribute their students into different skill groups. Their rankings corresponded well with verbal and memory tests and actual text recall, in that students who the teachers believed were high performers also had the highest scores on the tests and the best recall (Baddeley, et. al., 1985; Jackson & McLelland, 1979). In the last two experiments verbal tests were used and these tests were used to distribute students into verbal skill groups.

Taken together, these four general features of the data clearly suggest that there exists logic and ecological validity in the present data. High achievers (i.e., high verbal skill, three-year students) are the better performers and they also rate their general reading and memory abilities as better compared to low achievers (i.e., low verbal skill, two-year students). High and poor performers have correctly been identified by their teachers and by the verbal and memory tests used in these studies.

9.3 What's new – combining metamemory and metacomprehension

The fact that the present thesis has chosen to *combine* metamemory and metacomprehension in the same study is a contribution to the domain of metacognition. This way, we have studied how students manage to control and evaluate both their memory and their comprehension of texts. Both these processes are required in a reading situation to make sure that a reader both remembers and understands what he or she reads. Lovett and Pillow (1995) suggested that although comprehension and memory processes are intertwined in everyday life, these are two separate mental processes with different end states. One reason why the present students could predict their text recall better than their comprehension could be that they found reading to remember a much more demanding task than reading to comprehend. Eriksson (2000) asked questions regarding how accurately the students thought their ratings were. As many as 60% were satisfied, that is, thought that their calibrations of comprehension were accurate, less than 50% were satisfied with their predictions of text recall. Carroll et al. (1997) found that their students made higher JOLs for semantically related material than material they had overlearned. However, the best retention was obtained for overlearned material. Obviously, the students' personal feelings were at odds with real outcomes. There could be several reasons for this: students are familiar with the text topics which make them feel certain that they have understood the texts (Schneider & Laurion, 1993). They may not spend enough time reading the text to be able to answer the questions correctly (Hallam & Francis, 1998). In turn, this could lead to an illusion-of-knowing (Glenberg & Epstein, 1987). The students in the studies presented here were also asked to describe their way of going through the experimental tasks. Their answers indicated that they spent more time and effort trying to remember the texts than trying to understand them. One student wrote *"The one I read to remember I read much more carefully"*. Furthermore, the general questions have repeatedly shown that students rate their general ability to remember text as lower compared to their reading fluency and reading comprehension abilities (Gillström & Rönnberg, 1995; Eriksson & Rönnberg, 2000; Eriksson, 2000). One important conclusion then, is that trying to remember requires more effort and attention resulting in accuracy of memory related ratings

(Maki, et al., 1990). Dunlosky (1998) presents a suppression-hypothesis of comprehension monitoring, based on the idea that there is a strong human need to bring order into information. Research has shown that students have problems detecting inconsistencies in texts for the reason that propositions contradicting earlier ones are suppressed. Furthermore, accurate monitoring requires an awareness of many factors that affect comprehension, such as forgetting, judgement of learning, topic familiarity, and text knowledge. Thus, being in control of your own comprehension is presumably a much harder task than what students usually expect. Confronted with a memory monitoring task it is already from the beginning regarded as much effort requiring resulting in better awareness of performance.

9.4 What's new – verbal skill

Another contribution of new knowledge refers to the verbal skill results. There are different views of whether verbal skill groups differ in their metacognitive ability (Maki & Berry, 1984; Pressley et al., 1987; Sinkavich, 1995). Most of these do, however, favor high achievers as those having a better metacognitive ability on the whole. The present thesis has shown that high performing students have excelled in performance but showed a lesser ability to monitor their performance. Students of lower skill levels made more accurate ratings of performance, especially as postdiction accuracy of text recall is concerned. Davou, Taylor and Worrall (1991) showed that beginners rely on general strategies and thinking ability whereas experts rely on retrieval and pattern recognition. In this vein, LaBerge and Samuel (1985) proposed that skilled readers are not aware of the subskills of reading anymore. It has further been argued that high performing students function at a more automatic level when they read texts, letting them deploy their attention elsewhere (Ackerman, 1990). Their performance level is high but their ratings become less accurate. According to Winnie and Hadwin (1998), students sometimes put on their "auto-pilot" when they are confronted with well-known tasks. In these situations, students solve their tasks with little or no attention due to extended experience, domain knowledge, or skill. Marcus, Cooper, and Sweller (1996) suggested that there is a clear relation between working memory capacity and comprehension. When processing of information is automatic, the demands on working memory is minimal and learning is easily achieved. According to Klauer (1992), experts finish their tasks rapidly with little effort or attention due to them having larger "chunks" of possible moves to make and also "deeper" structures of knowledge. Logan (1988) concluded that automatic processing is fast, effortless, autonomous, stereotypical and more or less unavailable to consciousness, resulting in poor memory of the process as such.

What is suggested is that in the present experimental situations metacognitive thinking was not always needed. Otero and Campione (1992) found a relation between measures of metacognitive comprehension monitoring ability and academic achievement but also that this relation decreased with grade levels. At higher grade levels, emphasis is placed on knowledge and cognitive skill, whereas metacognitive skills seem to have little influence. At a certain point, a persons' subjective thinking of how to complete a task seems to be carried out very rapidly, as if it requires no conscious thought. At this stage, the tasks have been carried out so many times that the person can deploy their attention elsewhere (Ackerman, 1990). To quote Hacker (1998):

> *"..along with the ideas of "active" and "conscious" monitoring, regulation and orchestration of thought processes is the possibility that thinking about one's thinking, through repeated use or overlearning, may become automatized and consequently nonconscious"* (page 7).

One question that is likely to follow a quotation like this is whether automatized thoughts are metacognitive or simply cognitive. Hacker (1998) claims that many researches have taken the standpoint that metacognition is reserved for the conscious reportable thoughts. The conclusion that I make is that high verbal skill students easily can read and recall texts, which results in high levels of performance but a lesser need to engage themselves in metacognitive thinking on how to solve the tasks (Otero & Campione, 1992).

9.5 What's new – studying preferences

The data concerning reading strategies have generated new knowledge with respect to students' reading and subsequent performance ratings. The purpose of Study II was to study how different levels of instructed personal involvement affected calibration and prediction accuracies (Gillström & Rönnberg, 1995). It was assumed that more involvement and active reading should result in better cognitive and metacognitive performance. This assumption, however was not supported as the students were affected differently by the same instructions. At first, it was argued that this result was received due to students' cognitive style (Gillström & Rönnberg, 1995; Riding & Sadler-Smith, 1992), but later studies showed that delay sometimes changed students' preferences (Eriksson & Rönnberg, 2000; Eriksson, 2000). Thus, study preference is more a question of strategy than cognitive styles (Riding & Sadler-Smith, 1992). It should be noted that study preferences have only been accurate for text recall and not for comprehension. Even after delays of one week or one month students can identify which instruction makes them recall the most (Eriksson & Rönnberg, 2000; Eriksson, 2000). Gillström and Rönnberg (1995) found that when students studied texts in their preferred way, recall performance was at its best and prediction accuracy at its lowest. This suggests the possibility that under optimal conditions, due to less effort requirements, students can direct their attention to other things. Magliano, Little, and Graesser (1993) investigated how reading instructions, that varied in depth of processing affected reading performance. Some students were to analyze letters in words, sounding syllables or the like (i.e., superficial) whereas others were to summarize, formulate questions and so forth (i.e., deep). A third group were not given any specific reading instructions. All of the groups made accurate ratings a result which was discussed in terms of a "transfer of appropriate processing hypothesis". That is, inappropriate instructions are disregarded and the students strive at studying text at a deeper level. This thesis widens the perspective as for instructional effects on performance to suggest that choice and effectiveness of instruction is a personal matter. What works for me does not necessarily work for you.

9.6 Three critical points

One aspect to discuss is whether students should make one global rating per text or several, to increase reliability. Pressley et al (1987) found that global ratings were necessary if students were to make accurate ratings of test preparedness, whereas others have found that reliability improves with multiple text-segment analyses (Maki & Swett, 1987). In different ways, I have made attempts to resemble everyday school-life. I believe that students make global ratings more often than segment by segment. *"Have I understood what I read or not", "will I remember or not"*. Pressley et al. (1987) claimed that the reader has to see the whole text before judgements of learning or the like can be made, thus, favoring global ratings. In the present thesis, both types of ratings have been used but the global ratings have dominated.

Earlier it was discussed whether or not the restricted measure of two questions could have affected the calibration accuracy results. Would the pattern attained be different had there been more questions? Weaver (1990) suggested that reliability increased with more questions, yet reliability remained rather low regardless of the number of questions. The questions being used in this thesis differentiated verbal skill levels sufficiently well, if they had not, it would have been more worrisome. As the number of questions increases more extended texts would be needed, which, in turn, could result in other methodological concerns. Would not comprehension for extended text materials put harder demands on memory?

It is important to raise the question of experimental control of the students taking part. No attempts were made to identify dyslexics or students with other diagnosed reading and/or writing disorders. Focus has been placed on verbal skill and post-hoc analyses, and teacher ratings have sorted out high from low verbal skill students. Those with lower verbal skills demonstrated better metacognitive awareness compared to high verbal skill students. Students with high verbal skills have consistently been the best performers in the present studies. Would the results have been different knowing about, for example, dyslexia? The data patterns being as they are, my answer is no. One reason is that there should be only a relatively limited number of students diagnosed with this disorder within the present samples (Gustafson, 2000). Second, reading consists of two main components – decoding of words and comprehension (Gough & Tunmer, 1986). 'Poor readers' is not a homogeneous group but show substantial variability on measures of these two components of reading. Regardless of the exact nature of their reading problems, the poor readers are expected to perform less well compared to normal readers (Gathercole, Willis, and Baddeley, 1991; Lundberg & Höien, 1989; Samuelsson, Gustafson, Rönnberg, 1998). This is also the case for our low verbal skill students. Furthermore, Lundberg, Frost, and Petersen (1988) results showed that dyslexic kindergartners' metalinguistic skills were improved by training programs consisting of metalinguistic games and exercises. Thus, suggesting that poor readers also can be taught metacognitive awareness. Again, this pattern resembles the one being found for low verbal skill students who demonstrated prediction and postdiction accuracy.

9.7 Main conclusions

It is quite clear that the present students could predict their text recall with some degree of accuracy. This is a very consistent feature of our data and is strongly supported by the data reviewed in Pressley and Schneider (1997). The fact that calibration of comprehension displays different results is also compatible with previous data (Lin & Zabrucky, 1998). From a methodological point of view, it could be a problem that the number of participants in the experiments have differed, and that the number of questions measuring comprehension were restricted to two. The limited number of questions did have sufficient discriminating power, though, and in many studies one or two questions have been standard when measuring calibration accuracy (cf. Glenberg & Epstein, 1987). Therefore, the present thesis also presents a set of theoretical reasons as to why the metacomprehension results have varied between studies. What people comprehend from the same texts varies, and it could be that everyone can not answer the same set of questions. To come around this problem future research could include free-report of comprehension. That is, instead of the same questions to everyone, each student should describe his or her comprehension of the text. According to Koriat and Goldsmith (1997) free-report increases participants' monitoring control.

Furthermore, *knowing that you know* could also be a more demanding task than what readers usually expect. It could be that students take it more lightly to read to comprehend than reading to remember. In this vein, Eriksson (2000) reported that the students found reading to remember a much harder and demanding task than they found reading to comprehend. Finally, there are also social aspects of comprehension to consider. Could it be that students include expectations of success or failure into their ratings? Are the students, by themselves and by others, expected to do well or poorly in reading tasks? These are important questions to address in the future.

In closing, the present thesis suggests that the concept of effort is important to consider in this line of research. First, one of the reasons why the present students demonstrated better metamemory than metacomprehension, could be associated with effort. When asked, the students revealed that they worked harder in trying to remember compared with trying to comprehend. Thus, prediction and postdiction accuracy occurred.

Second, the lower verbal skill students made more accurate ratings of memory performance (especially postdictions) than did the high verbal skill students. This result suggests that the former group is in a position where metacognitive decisions are crucial (and effort demanding) and of most help in remembering texts.

Third, the preference hypothesis shows that across instructions and verbal skill, the students prefer different ways to study texts. These preferences interact with text recall and prediction accuracy but not with answering performance or calibration accuracy. Preferred study techniques do not demand as much effort as non-preferred study techniques, thus resulting in best possible text recall but reduced prediction accuracy.

The general assumption of this thesis is that metacognition is an important aspect of learning. We have an ability to think forwards on how we will perform but thinking backwards is more accurate. Evidently, there are many factors that affect thinking forwards and backwards and this thesis suggests that effort is one of them.

9.8 Future directions

Where will I go from here? There are many aspects of metacognition and learning that would be interesting to continue studying. Below a few of these ideas are presented.

- Effort and attention seem to be key elements that need further investigation. It would be productive to combine both of these with motivation to learn, with the ultimate purpose to create inspiring learning situations that the student is in control of and that brings him or her a longer way than traditional learning situations do. Also, the worlds of theory and practice need to meet, which an applied project of this sort could pave the way for.
- What if students were given a chance to create their own optimal learning situation? Choose their own type of text, instruction, reading time, and so forth. Can they do it? Would an individually specified learning situation be the best possible learning situation for a certain individual?
- A longitudinal study with the purpose to teach students to think cognitively and metacognitively about the learning processes. This study should involve students of different age groups. A study like this should be carried out in schools together with teachers. It should include efforts to increase students' self-esteem, self-awareness, and self-control.
- Metacomprehension and metamemory need to be investigated further. It could be argued, that in a learning situation students are first concerned with remembering, and as time goes by they turn into comprehenders. This is an aspect that should be considered in further detail, e. g., as

concerns where, when, and how students alter their behavior (cf. Conway, et al., 1997). Also, different ways to measure metacomprehension need to be employed. In this study, short expository texts were used followed by two questions, which could have had a negative effect on data. What is understanding of text and when does it occur?

- We no live in the 21th century. I think it is time to let students be in charge of their own learning. Students should be active learners in control, who can ask for the right learning material, teaching methods, and aids. Maybe a metacognitive approach is one way to accomplish this.

10. References

Ackerman, P. L. (1990). A correlational analysis of skill specificity: learning, abilities, and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 883-901.

Ackerman, P. L., Kanfer, R., & Goff, M. (1995). Cognitive and non-cognitive determinants and consequences of complex skill acquisition. *Report by Department of Psychology, University of Minnesota*. USA. February, 1995.

Allwood, C. M., & Granhag, P. A. (1996). Considering the knowledge you have: effects on realism in confidence judgements. *European Journal of Cognitive Psychology, 8*, 235-256.

Anderson, R. C., & Freebody, P. (1985). Vocabulary knowledge. In H. Singer & R. B. Ruddell (Eds.), Theoretical models and processes of reading, third edition. Newark: IRA.

Baddeley, A. D. (1999). *Essentials of human memory*. East Sussex: Psychology Press Ltd.

Baddeley, A. D., Logie, R., Nimmo-smith, I & Brereton, N. (1985). Components of fluent reading. *Journal of Memory and Language, 24*, 119-131.

Baker, L., & Brown, A. (1984). Metacognitive skills and reading. In P. D. Pearson, M. Kamil, R. Barr & P. Rosenthal (eds.), *Handbook of reading research*. New York: Academic Press.

Begg, I. M., Martin, L. A., & Needham, D. R. (1992). Memory monitoring: how useful is self-knowledge about memory. *European Journal of cognitive psychology, 4*, 195-218.

Begg, I., Duft, S., LaLonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language, 28*, 610-632.

Benjamin, A. S., Bjork, R. A., & Bennett, L. S. (1998). The mismeasure of memory: when retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*, 55-68.

Bouffard, T., Markovits, H., Vezeau, C., Boisvert, M., & Dumas, C. (1998). The relationship between accuracy of self-perception and cognitive development. *British Journal of Educational Psychology, 68*, 321-330.

Bouffard, T., Boisvert, M., Vezeau, C., & Larouche, C. (1995). The impact of goal orientation on self-regulation and performance among college students. *British Journal of Educational Psychology, 65*, 317-329.

Bouffard, T., Vezeau, C., & Bordeleau, L. (1998). The developmental study of the relation between combined learning and performance goals and students' self-regulated learning. *British Journal of Educational Psychology, 68*, 309-319.

Bristow, J., Cowley, P., & Daines, B. (1998). *Memory and learning. A practical guide for teachers*. London: David Fulton Publishers.

Brown, A. L., & Campione, J. C. (1978). The effect of knowledge and experience on the formation of retrieval plans for studying from texts. In Gruneberg, Morris & Sykes (Eds.), *Practical aspects of memory*. London: Academic press.

Carroll, M., & Korunika, S. (1999). The effect of text coherence and modality on metamemory judgements. *Memory, 7*, 309-322.

Carroll, M., & Nelson, T. O. (1993). Effects of overlearning on the feeling of knowing is more detectable in within-subject than in between-subject designs. *American Journal of Psychology, 106*, 227-235.

Carroll, M., & Nelson, T. O., Kirwan, A. (1997). Tradeoff of semantic relatedness and degree of overlearning: differential effects on metamemory and long-term retention. *Acta Psychologica, 95*, 239-253.

Conway, M. A., Gardiner, J. M., Perfect, T. J. Anderson, S. J., & Cohen, G. M. (1997). Changes in memory awareness during learning: the acquisition of knowledge by psychology undergraduates. *Journal of Experimental Psychology: General, 126*, 393-413.

Cull, W. L., & Zechmeister, E. B. (1994). The learning ability paradox in adult metamemory research: where are the metamemory differences between good and poor learners? *Memory & Cognition, 22*, 249-257.

Davidson, J. E., & Sternberg, R. J. (1998). Smart problem solving: how metacognition helps. In D. J. Hacker, J. Dunlosky, & Graesser, A. C. (Eds.), *Metacognition in educational theory and practice*. New Jersey: Lawrence Erlbaum Associates.

Davies, J., & Brember, I. (1999). Self-esteem and national tests in years 2 and 6: a 4-year longitudinal study. *Educational Psychology, 19*, 337-345.

Davou, 'B., Taylor, F., & Worrall, N. (1991). The interplay of knowledge and abilities in the processing of text. *British Journal of Psychology, 61*, 310-318.

Dixon, R. A., & Hultsch, D. F. (1983). Structure and development of metamemory in adulthood. *Journal of Gerontology, 38*, 682-688.

Dominoski, R. L. (1998). Verbalizing and problem solving. In D. J. Hacker, J. Dunlosky, & Graesser, A. C. (Eds.), *Metacognition in educational theory and practice* (pp. 25-45). New Jersey: Lawrence Erlbaum Associates

Dunlosky, J. (1998). Epilogue. Linking metacognitive theory to education. In D. J. Hacker, J. Dunlosky, & Graesser, A. C. (Eds.), *Metacognition in educational theory and practice* (pp. 367-381). New Jersey: Lawrence Erlbaum Associates.

Eriksson, Å. (2000). Comprehension calibration and memory prediction accuracy of texts: the effect of reading to remember and reading to comprehend instructions on performance and accuracy of ratings. Submitted manuscript.

Eriksson , Å. & Rönnberg, J. (2000). Comprehension calibration and memory prediction accuracy of texts: The effect of time, effort and preference on students' executive monitoring processes. submitted manuscript.

Flavell, J. H. (1979). Metacognition and cognitive monitoring. A new area of cognitive–developmental inquiry. *American Psychologist, 34*, 906-911.

Forrester-Pressley, D. L., & Waller, T.G. (1984). *Cognition, metareading, and reading*. New York: Springer-Verlag Inc.

Garner, R. (1987). *Metacognition and reading comprehension*. New Jersey: Ablex publishing corporation.

Gillström, Å. & Rönnberg, J. (1994). Memory prediction accuracy of text recall: Ease, effort and familiarity. *Scandinavian Journal of Psychology, 35*, 367-386.

Gillström, Å. & Rönnberg, J. (1995). Comprehension calibration and memory prediction accuracy of texts: Reading skill, reading strategies and effort. *Journal of Educational Psychology, 87*, 545-558.

Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition, 15*, 84-93.

Graham, S., Harris, K. R., & Troia, G. A. (1998). Writing and self-regulation: cases from the self-regulated strategy development model. In D. Chunk & B. Zimmerman (Eds.), *Self-regulated learning: From teaching to self-reflective practices* (pp. 20-41). New York: Guilford.

Graesser, A.C., Woll, S. B., Kowalski, D. J., & Smith, D. A. (1980). Memory for typical and atypical actions in scripted activities. *Journal of Psychology: Human learning and memory, 6*, 503-515.

Gutrie, J. T., & Kirsch, I. S. (1987). Distinctions between reading comprehension and locating information in text. *Journal of Educational Psychology, 79*, 220-227.

Hacker, D. J. (1998). Definitions and empirical foundations. In D. J. Hacker, J. Dunlosky, & Graesser, A. C. (Eds.), *Metacognition in educational theory and practice*. New Jersey: Lawrence Erlbaum Associates.

Hallam, S., & Francis, H. (1998). Is my understanding yours? A study of higher education students' reading for understanding and the effects of different texts. *Learning and Instruction, 8*, 83-95.

Helmke, A., & van Aken, M A. G. (1995). The causal ordering of academic achievement and self-concept of ability during elementary school: a longitudinal study. *Journal of Educational Psychology, 87*, 624-637.

Jackson, M. D., & McLelland, J. L. (1979). Processing determinants of reading speed. *Journal of Experimental Psychology: General, 108*, 151-181.

Johnson, R. E. (1970). Recall of prose as a function of the structural importance of the linguistic units. *Journal of Verbal Learning & Verbal Behavior, 9*, 12-20.

Karabenick, S. A. (1996). Social influences on metacognition: Effects of colearner questioning on comprehension monitoring. *Journal of Educational Psychology, 88*, 689-703.

Klauer, K. J. (1992). Teaching inductive thinking to highly able children. *European Journal for High Ability, 3*, 164-180.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General, 124*, 311-333.

Koriat, A. (1998) Illusions of knowing: the link between knowledge and metaknowledge. In Vincent Y. Yzerbyt, Guy Lories and Benoit Dardenne (Eds.), *Metacognition Cognitive and social dimension*. London: Sage publications.

Koriat, A., & Goldsmith, M. (1996a). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490-517.

Koriat, A., & Goldsmith, M. (1996b). Memory metaphors and the real-life/laboratory controversy: correspondence versus storehouse conceptions of memory. *Behavioral and Brain Sciences, 19*, 167-228.

Koriat, A., & Goldsmith, M. (1997). The role of metacognitive processes in the regulation of memory performance. Chapter to appear in G. Mazzoni & T. O. Nelson' s (Eds.), *Monitoring and control processes in metacognition and neuropsychology*. IIPDM Report No. 144, October 1997.

LaBerge, D., & Samuels, S. J. (1985). Toward a theory of automatic information processing reading. In Singer H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading*, third edition. Newark, DE: IRA Inc.

Lee, A. Y. L., Hutchinson, L. (1998). Improving learning from examples through reflection. *Journal of Experimental Psychology: Applied, 4*, 187-210.

Legree, P.J., Pifer, M.E., & Grafton, F.C. (1996). Correlations among cognitive abilities are lower for higher ability groups. *Intelligence, 23*, 45-57.

Lin L-M., & Zabrucky, K. M. (1998). Calibration of comprehension: research and implications for education and instruction. *Contemporary Educational Psychology, 23* 345-391.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492-527.

Lories, G., Dardenne, B., & Yzerbyt, V. Y. (1998). From social cognition and metacognition. In Vincent Y. Yzerbyt, Guy Lories and Benoit Dardenne (Eds.), *Metacognition Cognitive and social dimension*. London: Sage publications.

Lovett, S. B., & Pillow, B. H. (1995). Development of the ability to distinguish between comprehension and memory: evidence from strategy- selection tasks. *Journal of Educational Psychology, 87*, 523-536.

Lovett, S. B., & Pillow, B. H. (1996). Development of the ability to distinguish between comprehension and memory: evidence from goal-state evaluation tasks. *Journal of Educational Psychology, 88*, 546-562.

Lovelace, E. A. (1984). Metamemory: monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 756-766.

Lundberg, I., Frost, J., & Petersen, O-P. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23, 263-284.*

Lundberg, I., & Höien, T. (1989). Phonemic deficits: A core symptom of developmental dyslexia? *The Irish Journal of Psychology, 10*, 579-592.

Magliano, J. P., Little, L. D., & Graesser, A. C. (1993). The impact of comprehension instruction on the calibration of comprehension. *Reading Research and Instruction, 32*, 49-63.

Maki, R. (1998). Test predictions over text material. In D. J. Hacker, J. Dunlosky, & Graesser, A. C. (Eds.), *Metacognition in educational theory and practice* (PP 117-144). New Jersey: Lawrence Erlbaum Associates.

Maki, R., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 663-679.

Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 609-616.

Maki R. H., & Serra, M. (1992). The basis of test predictions for text material. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18*, 116-126.

Maki, R., & Swett, S. (1987). Metamemory for narrative text. *Memory and Cognition, 15*, 72-83.

Marcus, N., Cooper, M., & Sweller, J. Understanding instructions. *Journal of Educational Psychology, 88*, 49-63.

Mazzoni, G., & Cornoldi, C. (1993). Strategies in study time allocation: Why is study time sometimes not effective? *Journal of Experimental Psychology, General, 122*, 47-60.

McDaniel, M. A., Einstein, G. O., Dunay, P. K., & Cobb, R. E. (1986). Encoding difficulty and memory: towards a unifying theory. *Journal of Memory and Language, 25*, 645-656.

McDaniel, M.A. (1984). The role of elaborative and schema processes in story memory. *Memory and Cognition, 12*, 46-51.

Nelson, T. O., Kruglanski, A.W., & Jost, J. T. (1998). Knowing thyself and others: progress in metacognitive social psychology. In Vincent Y. Yzerbyt, Guy Lories and

Benoit Dardenne (Eds.), *Metacognition Cognitive and social dimension*. London: Sage publications.

O´Brien, E. J., & Meyers, J. L. (1985). When comprehension difficulty improves memory for text. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 12-21.

Otero, J. (1998). Influence of knowledge activation and context on comprehension monitoring of science texts. In D. J. Hacker, J. Dunlosky, & Graesser, A. C. (Eds.), *Metacognition in educational theory and practice* (pp. 145-164). New Jersey: Lawrence Erlbaum Associates.

Otero, J., & Campanario, J. M. (1992). The relationship between academic achievement and metacognitive comprehension monitoring ability of Spanish secondary school students. Educational and Psychological Measurement, 52, 419-430.

Otero, J., & Campione, J. M. (1990). Comprehension evaluation and regulation in learning from science texts. *Journal of Research in Science Teaching, 27*, 447-460.

Paris, S. G., & Meyers, II, M. (1981). Comprehension monitoring, memory and study strategies of good and poor readers. *Journal of Reading Behavior, 13*, 5-22.

Persson, U-B. (1994). *Reading for understanding. An empirical contribution to the metacognition of reading comprehension*. Linköping Studies in Education and Psychology No 41. Linköping University, Department of Education and Psychology.

Powel, W. D., & Gray, R. (1995). Improving performance predictions by collaboration with peers and rewarding accuracy. *Child Study Journal, 25*, 141-154.

Pramling, I. (1987). *Vad är metakognition*? Publikation från institutionen för pedagogik, Göteborgs universitet, nr 7.

Pressley, M., & Schneider, W. (1997). *Introduction to memory development during childhood and adolescence*. London: Lawrence Erlbaum Associates.

Pressley, M., Snyder, B. L., Levin, J. R., Murray, H. G., & Ghatala, E. S. (1987). Perceived readiness for examination performance (PREP) produced by initial reading of text and text containing adjunct questions. *Reading Research Quarterly, 12*, 219-236.

Riding, R., & Sadler-Smith, E. (1992). Type of instructional material, cognitive style and learning performance. *Educational Studies, 18*, 323-339.

Rueda, R., & Mehan, H. (1986). Metacognition and passing: Strategic interaction in the lives of students with learning disabilities. *Anthropology & Education Quarterly, 17*, 145-165.

Schommer, M., & Surber, J. R. (1986). Comprehension-monitoring failure in skilled adult readers. *Journal of Educational Psychology, 78*, 353-357.

Schneider, W. (1985). *Developmental trends in the metamemory – memory behavior. Relationship: an integrative review.* In D. L. Forrester-Pressley, G. E. MacKinnon, & T. G.Waller (Eds.), Cognition, Metacognition, and Human Performance, vol. 1. New York: Academic press.

Schneider, S. L., & Laurion, S. K. (1993). Do we know what we've learned from listening to the news? *Memory & Cognition, 21*, 198-209.

Simpson, S. M., Licht, B. G., Wagner, R. K., & Stader, S. R. (1996). Organization of children's academic ability-related self-perceptions. *Journal of Educational Psychology, 88*, 387-396.

Sinkavich, F. J. (1995). Performance and metamemory: do students know what they don't know? *Journal of Instructional Psychology, 22*, 77-88.

Skolverket. (1996). Grunden för fortsatt lärande. SKOLV-R-115-SE.

Spiro, R. J., & Myers, A. (1984). Individual differences and underlying cognitive processes in reading. In P. D. Person (Ed), *Handbook of reading research*. New York: Longman Inc.

Stark, R., Renkl, A., Gruber, H., & Mandl, H. (1998). Indeed, sometimes knowledge does not help: A replication study. *Instructional Science, 00*, 1-17.

Stipek, D., & Gralinski, J. H. (1996). Children's beliefs about intelligence and school performance. *Journal of Educational Psychology, 88*, 397-407.

Torrence, M., Thomas, G.V., & Robinson, E. J. (1993). Training in thesis writing: an evaluation of three conceptual orientations. *British Journal of Educational Psychology, 63*, 170-184.

Wade, S. E., & Trathen, W. (1989). Effects on self- SELECTED study methods on learning. *Journal of Educational Psychology, 81*, 40-47.

Weaver, C. A. III (1990). Constraining factors in calibration of comprehension. *Journal of Experimental Psychology: Learning, memory and cognition, 16*, 214-222.

Wenestam, C-G. (1993). A critique of research on cognition and cognitive processes. *British Journal of Educational Psychology, 63*, 34-45.

Winne, P. H., & Hadwin, A. F. (1998). Studying self-regulation. In D. J. Hacker, J. Dunlosky, & Graesser, A. C. (Eds.), *Metacognition in educational theory and practice* (pp. 277-304). New Jersey: Lawrence Erlbaum Associates.

Zimmerman, B. J. (2000). Self-efficacy: an essential motive to learn. *Contemporary Educational Psychology, 25*, 82-91.

Studie I

Scandinavian Journal of Psychology, 1994, 35, 367–386

Prediction accuracy of text recall: Ease, effort and familiarity

ÅSA GILLSTRÖM and JERKER RÖNNBERG

Department of Education and Psychology, Linköping University, Linköping, Sweden

Gillström, Å. & Rönnberg, (1994). J. Prediction accuracy of text recall: Ease, effort and familiarity. *Scandinavian Journal of Psychology, 35,* 367–386.

Prediction accuracy of text recall was studied in two experiments. Text characteristics (i.e., consistency and distinctiveness) were manipulated in Experiment 1, and familiarity with the reading-task in Experiment 2. The results were also analyzed and discussed in terms of easy processing (Experiment 1), and in terms of increased and more active processing (Experiment 2). Text characteristics did not affect prediction accuracy. However, being familiar with the reading-task led to good and long-lasting prediction accuracy. Thus, subjects reading a school-book text, instructed to learn the contents of it demonstrated reliable memory awareness, both for immediate recall and for delay of one week. It was also suggested that increased processing demands and active reading enhances prediction accuracy.

Key words: Prediction accuracy, text recall, instructions, ease, effort, familiarity.

Åsa Gillström, Department of Education and Psychology, Linköping University, S-581 83 Linköping, Sweden.

Readers who realize that they have not completely understood a text passage and take appropriate actions to improve their understanding, show metacognitive skill. We do not always, however, behave this maturely (Pressley *et al.*, 1987). Still, studies of metacognition and metamemory have offered new study and teaching techniques which emphasize the necessity of more analytic reading approaches to increase students' awareness of what has been understood and/or remembered (Costa, 1984; Cross & Paris, 1988; Finley & Seaton, 1987). There is some evidence that we are able to decide what will and will not be remembered, with the restrictions that this personal knowledge is *age-related* (Dixon & Hultsch, 1983; Pullyblank *et al.*, 1985; Rabinowitz *et al.*, 1982; Suzuki-Slakter, 1988; Bird, 1979), *skill-dependent* (Byrd & Gholson, 1985; Dermody, 1988; Sinkavich, 1988), and *task-dependent* (Begg *et al.*, 1989; Epstein *et al.*, 1984; Weaver III, 1990).

Skill-dependence is a complicated issue. It has been clearly demonstrated that high achieving students are better performers than low achieving students (e.g. Paris & Meyers, 1981). High achievers are also better incidental learners than low achievers (e.g. Necka *et al.*, 1992). Haneggi and Perfetti (1992) found that reading ability was more predictive for comprehension than were different types of text processing. But is it necessarily so, that high achievers know more about their internal processes than low achievers? Unfortunately, the results are ambiguous.

Maki and Swett (1987) found a negative relationship between achievement level and prediction accuracy. Pressley *et al.* (1987) found no evidence that high achievers showed better prediction accuracy than low achievers. Maki and Berry (1984) found that high achievers more accurately calibrated comprehension than low achievers. Pressley *et al.* (1987) suggested that the difference between their study and the one by Maki and Berry (1984) was that of global calibration compared to smaller text-part predictions. According to Pressley *et al.* (1987), students should be advised to make global estimates of comprehension and informed of the necessity of studying still more. Calibrations should be made after reading

© 1994 Scandinavian University Press. ISSN 0036-5564

and/or after testing rather than before reading. In the present study, the subjects varied in achievement level. Also, the subjects in Experiment 1 predicted recall performance for each paragraph in the text passages, whereas in Experiment 2, the subjects made global predictions of recall performance. In both experiments prediction ratings were made after reading.

Task-dependence of metamemory is perhaps the most complicated issue. As Maki and Serra (1992) conclude, it is quite clear that people to a large extent can predict their recall of lists of words and/or sentences. The degree of prediction accuracy is even increased when subjects are given extra study time or an opportunity to prior study (Lovelace, 1984; Thompson & Barnett, 1985). Begg *et al.* (1992) found that cued-review, in which one of the words in a word-pair was reviewed (e.g. railroad-?) increased prediction accuracy, compared to pair-review (e.g. railroad-mother), in which the whole pair was reviewed. One reason was that cued-review led to self-evaluation. If the subject was unable to remember the target-word at review, he/she could accurately assume that this word would not be remembered later on. In contrast to Lovelace (1984) and Thompson & Barnett (1985), Begg *et al.* (1992) found no evidence that review improved prediction accuracy. In a previous study, Begg *et al.* (1989) found that accurate memory predictions of words require that the same processes are used both for predictions as well as for tests. Begg *et al.* (1989) also suggested that prediction ratings are implicit judgements of how easily items are processed while predicting. Prediction accuracy is substantial if the factors that cause easy processing also lead to successful remembering. In Experiment 1, the suggestions made by Begg *et al.* (1989) were tested, in that the text material, the instructions etc. should be easy to process, and therefore should lead to better memory awareness and recall.

The present study is concerned with prediction accuracy of text. Both memory monitoring and comprehension monitoring of texts (calibration) are often related since both are related to memory (Pressley *et al.*, 1987). If something is understood it is usually, but not always, retrievable from memory, and vice versa. Maki and her colleagues seem rather optimistic about students' comprehension monitoring, that is, students know to some extent how well they will perform on subsequent tests (Maki *et al.*, 1990; Maki & Sera 1992), whereas Glenberg and his colleagues report more pessimistic results; the students seem to be unable to judge their performance on subsequent tests (Glenberg *et al.*, 1982; Epstein *et al.*, 1984; Glenberg *et al.*, 1987).

Both experiments in the present study investigated subjects' ability to predict their recall performance of texts. In Experiment 1, text characteristics, such as distinctiveness and consistency, were manipulated. Experiment 1 was also conceived of as a study of ease of processing, in that the text material used was written in a very simplified form and should be easy to comprehend. Begg *et al.* (1989) suggested that ease of processing is one important aspect of accurate prediction of memory performance, at least for words. This experimental design thus led to a testing of whether or not the ease of processing hypothesis holds true for prose as well. The texts were presented on a computer screen, paragraph by paragraph. For each paragraph the subjects predicted their recall performance.

In Experiment 2, familiarity with the reading-task was manipulated. Subjects either read a familiar type of text (i.e., school-book text) with a familiar instruction (i.e., to learn), or a less familiar type of text (i.e., fairy-tale) with a less familiar instruction (i.e., to teach). A third group of subjects read a school-book text with teaching instruction, and a final fourth group, a fairy-tale with learning instruction. The text material used in Experiment 2 was much longer than the prose passages used in the first experiment. The subjects formed their own key for recall, in that they underlined what they believed were important information based on their instruction. It was therefore assumed that the encoding demands were increased. Hence, the subjects had to put more effort in and be more active as readers, compared to Experiment 1. Finally, the texts were presented to the subjects on paper. During 15 minutes the subjects could read, study and make underlinings according to instruction. After reading the texts, the subjects made one global prediction rating of their recall performance.

Increased encoding demands necessitate the processing of many associations in order to comprehend (McDaniel *et al.*, 1989). With less demanding tasks the identification of words is more automatic or subconscious, but with increased demands more conscious routines have to be activated to identify words and meaning (McDaniel, 1984). McDaniel (1984) used deleted-letter manipulations, which increased recall performance. Maki *et al.* (1990), replicated that filling in deleted letters in a text led to better recall performance, and more distinct memory of these deleted ideas. More importantly, they also found that the calibrations of comprehension became more accurate with increased processing.

To sum up: In the present study text characteristics and familiarity aspects on prediction accuracy of text recall were manipulated. However, based on the fact that previous research has demonstrated inconsistent results regarding this memory knowledge, the present data were interpreted and discussed in terms of ease of processing (Experiment 1), and increased and more active processing (Experiment 2). The procedural differences between the experiments were also discussed.

EXPERIMENT 1

Maki and Swett (1987) studied different forms of text characteristics and found that recall performance and prediction accuracy were better for the inconsistent text passages compared to consistent text passages, hence suggesting a von Restorff-effect.

In Experiment 1, four short prose passages (Appendix 1 and 2) were used. The prose passages varied in consistency (Maki & Swett, 1987), but also in distinctiveness, based on the assumption that distinct items are more easily remembered and retrieved from memory (Eysenck & Eysenck, 1980). Hence, would the combination of inconsistency (Maki & Swett, 1987) and distinctiveness (Eysenck & Eysenck, 1980) receive even higher recall performance, and more importantly, more reliable predictions, than inconsistent text passages alone?

Between prediction ratings and recall the subjects took two verbal tests (Antonyms/Synonyms and Analogies) and two memory tests (Reading span, tapping working memory; and Lexical access speed, tapping long-term memory). The main reason for using the tests was to study if these could predict recall performance better than the subjective ratings. Based on previous research it was hypothesized that good text recall performance requires good working memory capacity and verbal ability (Baddeley *et al.*, 1985). The lexical access speed test was included because such a test has been shown to be a good predictor of reading comprehension (Jackson & McLelland, 1979). It was also hypothesized that the high achievers' performance should be better than the low achievers', since the tests are related to reading ability. In addition, these tests (for simplicity named objective tests) made it possible to compare the subjective ratings of recall performance with the objective tests as potential predictors of recall performance.

Experiment 1 was initially designed to study the effect of text characteristics on prediction accuracy of text recall (Maki & Swett, 1987). As argued, the experiment can also be interpreted as a situation in which the text processing characteristics of the task were easy to handle. The role of the readers was also rather passive in that no active text processing was required. In this way, it could no be evaluated whether easy processing is conducive to prediction accuracy or not (Begg *et al.*, 1989).

Method

Subjects. A total of 80 subjects participated in Experiment 1. They were all pupils in the 9th grade in the Swedish compulsory school. Nine Swedish teachers, from two different schools, were instructed to select those of their pupils they rated had a good as well as a poor reading comprehension ability. The selected subjects were then divided into one group of 40 high achievers, and one group of 40 low achievers. The subjects were tested individually. The experiment lasted for about an hour for each subject, and 40 SEK was recieved for participation.

General design and procedure. Four stories were written for the experiment. Each story consisted of approximately 150 words, divided into 10 paragraphs (see Appendix 1 and 2). Consistency and distinctiveness of the stories were orthogonally manipulated within these four stories. The stories were named C/D (consistent/distinctive), CNd (consistent/nondistinctive), IcD (inconsistent/distinctive), and finally, IcNd (inconsistent/nondistinctive). CD and IcD were basically the same story, with the exception of consistency, which was manipulated by changing one paragraph into being inconsistent with the rest of the theme in IcD. The same relationship existed between CNd and IcNd, in that consistency was manipulated by changing one paragraph into being inconsistent with the rest of the theme in IcNd. The distinctiveness in CD and IcD was manipulated in paragraph 7 (dramatic event).

The texts were presented by means of an Apple computer (Lisa), and the 10 paragraphs were programmed to be displayed, one at a time, on the computer screen. The time it took the subject to read and make a prediction rating of each paragraph was registered by the computer programme TIPS (Ausmeel, 1988). The subjects were instructed to press a predefined button on the computer to start the presentation of the next paragraph, thus setting the response time for each paragraph. The maximum time that the paragraph was displayed on the screen was two minutes. After this interval, the next paragraph was automatically presented on the screen. For each subject the total average reading time was calculated and used as an objective measure together with the other verbal and memory tests.

Each subject was randomly assigned to one of the four experimental texts. The subject was instructed to read a paragraph through and then to make a prediction rating of how much he/she would be able to recall of the paragraph after a delay of one hour. For each paragraph a 7-point rating scale was used, ranging from "*Will probably not be able to recall anything*" (1) to "*Will probably be able to recall everything*" (7). This reading–prediction procedure was the same for all 10 paragraphs. To ensure that the subjects had understood the experimental intentions correctly, and to make them accustomed to the experimental situation, the subjects were given a short practice text prior to the experimental test session.

Before text recall the subjects made ratings in response to three text judgement questions concerning their opinion of text difficulty, distinctiveness and consistency. For each question a 7-point rating scale was used. The questions were as follows: "*Rate how difficult the text was to comprehend*", the scale ranged from "*Not difficult at all*" (1) to "*Very difficult*" (7); "*To what extent did you find distinctive elements in the text*", the scale ranged from "*To no extent at all*" (1) to "*To a large extent*" (7); "*To what extent did you find the text logical and connected*", the scale ranged from "*Not logical and connected*" (1) to "*Completely logical and connected*" (7). Thereafter, the subjects were given the memory and verbal tests. Finally, there was no time-limit for text recall. The subjects were asked to recall as much as possible, in a verbatim fashion.

Objective tests. The subjects were given the objective tests between prediction rating and recall. The tests were given in two orders, either the subject took the two verbal tests first and then the memory tests, or vice versa. These test orders were counterbalanced across text type and achievement level.

Verbal tests. The verbal tests chosen assessed the subjects' word knowledge and verbal inductive ability (Westrin, undated). Each verbal test was presented to the subjects on printed sheets. For each test there were two practice examples and the instructions were given orally to the subjects. The first test was an Analogy test with 27 tasks (the subjects were given $5\frac{1}{2}$ minutes to complete the test). The second test was an Synonym/Antonym test with 29 tasks (the subjects were given 5 minutes to complete the test). The measure used was the total number of correct answers on each of the tests (i.e., maximum 27 or 29 points).

Memory tests. The memory tests were administered by the TIPS-programme (Ausmeel, 1988) on the Apple computer (for detailed description, see Rönnberg *et al.*, 1989). One of these tests tapped *working-memory* capacity. The subjects were asked to read a sentence, presented word by word on the computer at a rate of one item per 0.8 sec. For each sentence, the subjects were to decide (by saying yes or no) whether the sentence was logical or not. Three to six sentences were presented this way during each trial (for a total amount of 54 sentences). After each trial, the subjects were asked to recall, in correct serial order, the last word of each of the set of sentences just presented. The measure used was the total number of last words correctly recalled in serial order. The other memory test tapped *lexical access speed*. The subjects were asked to decide whether a string of letters (3 letters) was a correct

Swedish word or not. Each string of letters (the total amount was 100 words; 50 correct Swedish words pl. 50 lures) was presented on the computer for a maximum of 2 seconds. The subjects were given practice examples of each test, and they were given oral instructions. The measure used was the latency, collapsed across yes/no responses.

Results

Overall text judgements. The mean value of the subjects' ratings of text difficulty, distinctiveness, and consistency was calculated. The overall level of rated difficulty was 40%[1] (high achievers: 39% and low achievers: 43%). A one-factor ANOVA on ratings of distinctiveness, with the type of text as a between subjects factor, showed that the distinct texts were regarded as more distinct (53%) than nondistinctive texts (37%), $F(1,76) = 7.6$, $p < 0.05$. Also, a second one-factor ANOVA on ratings of consistency, with type of text between subjects factor, showed that the consistent texts were regarded as more consistent (84%) than inconsistent texts (70%), $F(1,76) = 11.9$, $p < 0.05$. The results of these text judgement questions confirmed that the prose passages had been easy to read, and that the subjects had observed the manipulated text characteristics.

Prediction ratings. A mean prediction rating value was calculated for each subject, based on the average of the subjects' 10 prediction ratings. No significant differences in prediction ratings between text types, $F(1,78) = 0.06$, $p = 0.80$, were obtained. The ratings ranged between 64% and 73%.

Recall. The data analyses were based on recall of propositions, since this scoring procedure is most commonly used in this line of work (Maki & Swett, 1987). The four prose passages were divided into 33 propositions by two independent raters (van Dijk & Kintsch, 1983). For each subject the mean number of recalled propositions was calculated, averaged over the 10 paragraphs. The scoring of propositions was done in a lenient fashion (i.e., synonyms, change of inflections, and change of singular and plural were accepted).

Mean number of recalled propositions for each text type, pooled over achievement level ($n = 20$) was for *CD*: 51%, *CNd*: 47%, *IcD*: 55%, and *IcNd*: 58%. A one factor ANOVA on recall performance with type of text as between subjects factor revealed no significant difference in recall performance between the text types, $F(3,76) = 1.25$, $p = 0.30$. A one factor ANOVA on recall performance, with achievement level as between subject factors, revealed that high achievers recalled (58%) significantly more than low achievers (47%), $F(1,78) = 7.9$, $p < 0.05$.

The idea that inconsistency leads to higher recall performance could not be confirmed (Maki & Swett, 1987). The means point in that direction but no significant results were attained. The fact that high achievers are better "recallers" than low achievers was replicated (e.g., Paris & Meyers, 1981). Recall performance was also scored on mean number of recalled content words. The same results as with recalled propositions were attained.

Prediction accuracy. Prediction accuracy, measured by correlating mean prediction ratings with mean recall of propositions, was close to zero at the overall level ($n = 80$), between high and low achievers ($n = 40$), between the different text types ($n = 20$), and for each text type and achievement level ($n = 10$). To study the effect of the manipulated distinctiveness and consistency the correlation coefficient was calculated between mean prediction and mean recall for the distinctive paragraph in the CD-condition, and for the nondistinctive paragraph in the CNd-condition. The same calculation was performed with the inconsistent paragraph in the IcD-condition and the consistent paragraph in the CD-condition, as well as the consistent paragraph in the CNd-condition and the inconsistent paragraph in the IcNd-condition. Neither pooled over achievement level, nor for high and low achievers respectively, were any significant results found.

---

[1] For all data-analyses, the ratio (%) between rating and maximum scale value (i.e., 7) was calculated.

Objective prediction accuracy. As Table 1 shows, the hypothesis that the high achievers should perform better than the low achievers on the objective tests was confirmed. No significant difference was found for the reading time variáble, which can be explained by the fact that all subjects, irrespective of achievement level, had regarded the texts as easy to read and only spent on average 24–25 sec to read the texts (see overall text judgements).

The main question was whether or not the objective tests could predict recall performance with accuracy. Therefore, correlation coefficients were calculated for the overall level ($n = 80$), and for the two reading levels ($n = 40$), respectively.

Overall, significant relations were found between working memory, verbal inductive ability, work knowledge and reading time, respectively, and mean recall of propositions (see Table 2). Some differences were found between the achievement levels. High achievers' word knowledge and working memory ability correlated singificantly with recall performance. For low achievers, both verbal inductive ability and word knowledge correlated significantly with recall performance, whereas the memory tests did not. For both groups of subjects reading time showed the most substantial correlation with recall performance. Thus, subjects who used more time to study the texts, also demonstrated better recall performance. Verbal ability seems to be a crucial factor for low-achieving students' recall performance. Therefore, a good working memory capacity is of minor importance if understanding of words and language is not there. Lexical access speed did not correlate with mean recall. Obviously, this type of long-term memory function (Rönnberg, 1990) is not a prerequisite for recall performance.

Table 1. *High and low achievers' mean performance on the objective tests ($n = 40$) in Experiment 1*

	Reading level	
Test	High achievers	Low achievers
Analogy test	16.975***	10.675
Antonym/Synonym test	16.175***	10.450
Working memory	28.250***	21.750
Lexical access speed	0.892**	0.998
Reading time	24.470	25.484

**$p < 0.01$, and ***$p < 0.001$ refers to the difference between high and low achievers, assesed by a t-test.

Table 2. *Correlation coefficients between mean recall of propositions and mean performance on the objective tests for all subjects ($n = 40$), in Experiment 1*

	Subjects		
Test	All	High achievers	Low achievers
Analogy test	0.41**	0.16	0.38*
Antonym/Synonym test	0.56**	0.34*	0.62**
Working memory	0.40**	0.39*	0.22
Lexical access speed	−0.15	−0.04	−0.08
Reading time	0.55**	0.43**	0.70**

*$p < 0.05$
**$p < 0.01$

Discussion

Confirming previous studies, high achievers demonstrated significantly better recall performance than low achievers (e.g., Paris & Meyers, 1981) Maki and Swett (1987) suggested that inconsistent text materials, as opposed to consistent, should lead to better recall performance and prediction accuracy. Experiment 1 could not replicate these results. Inconsistent texts neither led to better recall performance nor to better prediction accuracy of text recall. On the whole, no prediction accuracy was found for any of the experimental conditions or achievement levels. In sharp contrast to the lack of subjective prediction accuracy, the objective tests, except for lexical access speed, predicted the subjects' recall performance with reliable accuracy. The objective tests also discriminated high from low achieving students.

The fact that the objective tests correlated with recall performance is not surprising. Necka *et al.* (1992) also found that high-ability subjects performed significantly better than low-ability subjects on verbal tests. They also found that high achievers are better incidental learners. Thus, incidental learning may be related to verbal ability. Working memory is also related to recall and comprehension (Haenggi & Perfetti, 1992; Baddeley *et al.*, 1985). Reading time showed the highest correlation coefficient with recall performance, both for high achievers ($r = 0.50$) and low achievers ($r = 0.70$). Mean prediction ratings of recall performance were not correlated with mean reading time, $r = -0.05$. Thus, this seems to suggest that memory, but not metamemory, gains from more reading time.

Experiment 1 did not replicate the levels of prediction accuracy reported by Maki and Swett (1987). The reason can be found in the number of idea units to be recalled. Weaver (1990) argued that calibrations of comprehension becomes more accurate and reliable when the number of test items per text is increased. This could also be an important prerequisite for recall of prose passages. Hence, a larger set of paragraphs to-be-recalled could yield an increase of both reliability and effort, which in its turn could lead to higher prediction accuracy. Maki and Swett (1987) used texts consisting of 180–250 words, divided into 42 propositions. The texts in Experiment 1 consisted of approximately 150 words, divided into 33 propositions. The lack of prediction accuracy could also be due to the nature of the rating question employed (Horgan, 1990). The subjects may have had problems in interpreting the meaning of the scale. For instance, how much recall performance does 5 on the scale represent? To correct for this potential problem, a more familiar scale was used in Experiment 2. Thus, the rating scale ranged from 1 to 100%.

The lack of prediction accuracy also prompted an examination of whether or not over- and/or underestimations[2] were the cause. A clear pattern of overestimations was found. Of all subjects, 49% overestimated their recall performance whereas only 11% underestimated. Among the 40 low achievers, 57% overestimated their recall performance, whereas only 5% underestimated. Among the 40 high achievers, 40% overestimated, and 17% underestimated their recall performance. For the different text types most overestimations were done for CNd: 65%, and the least for IcNd: 30%. In between, CD: 55% and IcD: 45%. These overestimations together with the fact that the subjects only spent on average 25 sec/paragraph suggests a poor knowledge about memory requirements. The subjects could have spent 2 minutes reading each paragraph, but they satisfied themselves with much less. What could have led to these overestimations? One suggestion is that it could be due to the easy demands and the less elaborative reading procedures imposed by the Experiment. The fact that the subjects made prediction ratings for each paragraph instead of one global prediction of the text could also have mislead the subjects in believing that they would be able to recall more than they actually did.

---

[2] Prediction ratings at least 11% higher than or under actual recall performance were regarded as over- and underestimations.

The alternative position is one in which prediction accuracy and memory performance is assumed to be based on increased effort requirements at the encoding stage (O'Brien & Meyers, 1985; McDaniel *et al.*, 1989). Thus, performance and knowledge about performance should increase as the ongoing processes go from automatic and subconscious to conscious and more deliberate processing (McDaniel, 1984), an empirical fact that Maki *et al.* (1990) was able to demonstrate. Their subjects' calibrations of comprehension were most accurate when they, as suggested by McDaniel (1984), filled in deleted letters in the text, as opposed to intact texts. Thus, increased and more active processing demands led to better calibration accuracy.

EXPERIMENT 2

The purpose of Experiment 2 was twofold. It was designed to study the effect of familiarity with a situation, and its relation to prediction accuracy. However, as Experiment 1 showed poor student knowledge of memory requirements, the putative effect of increased processing also seemed important to take into consideration (McDaniel *et al.*, 1989; Maki *et al.*, 1990; Maki & Serra, 1992). In addition, it seemed important to let the subjects make global prediction ratings instead of bit by bit predictions. To achieve an increase of encoding demands, and induce more elaborative reading procedures, a different approach was chosen than in Experiment 1. In the first place, all subjects read text materials (see Appendix 3 and 4) which were considerably extended (Weaver, 1990). Furthermore, all subjects formed their own key for recall by underlining important words and/or sentences in the texts (Greenwald & Banaji, 1989). Finally, all subjects were tested a second time, after a week's delay.

Levels of familiarity with a situation was manipulated between the subjects. To that end, the subjects either read a school-book text (i.e., descriptive exposition) or a fairy-tale (i.e., narration; McDaniel *et al.*, 1986), and were at the same time instructed to either assume the role of learner or teacher. It was assumed that teaching as opposed to learning would involve a less frequent and different study behavior. The subjects given teaching instructions would have to consider what other people need to be informed about in order to understand the contents. Persson (1990) found that 5th and 8th graders found school-book texts more familiar than other types of text (including fairy-tales). According to her subjects, the school-book text was also associated with more demands because of its clear relation to learning.

Thus, the text materials were combined with instructions, with the most familiar situation being a learner of a school-book text, and the least familiar situation being a teacher of a fairy-tale.

Method

Subjects. A total of 129 subjects participated in the study. They were all pupils in the 9th grade of the Swedish compulsory school. As in Experiment 1, teacher ratings of reading comprehension ability was used to divide the subjects into three achievement levels: *high achievers* (44 subjects), *normal achievers* (49 subjects) or *low achievers* (36 subjects).

General design and material. All subjects were tested twice: Immediately, and after a week's delay. The immediate test session took about an hour, and the delayed session about half an hour. The experiment took place in ordinary classrooms and 15–20 subjects participated at the same time.

Two different types of texts were used in Experiment 2. The school-book text was a history book text (Kahnberg & Lindeberg, 1964) about the lives of the Swedish Vikings (Appendix 4). The text can be defined as a descriptive exposition or as an instructional text. The fairy-tale was a folktale about a chased hare (Förlaget Barrikaden, Stockholm, 1980). This text can be described as narration or fiction

(Appendix 3). The texts consisted of approximately 680 words each. Each subject read only one of these texts.

Two types of instruction were used in the experiment. Either the subjects were to assume the role of a learner, that is, read the text to *learn* the contents of it, or the subjects were to assume the role of a teacher, that is, read the text to *teach* somebody else the contents of it. Common to both instructional conditions was that the subjects had to underline the words and/or sentences in the texts they found to be important on the basis of the instruction given. A minimum of 10 words and/or sentences had to be underlined; there was no upper limit. Each subject was given only one of the two instructions.

Thus, four main experimental conditions were created by combining type of text with type of instruction: 1) *L*earning a *S*chool-book text (LS) 2) *T*eaching a *S*chool-book text (TS) 3) *L*earning a *F*airy-tale (LF) 4) *T*eaching a *F*airy-tale (TF). Due to the nature of classroom experiments with groups of subjects, the random assignment of subjects to conditions produced somewhat unequal *n*'s: LS—29 subjects, TS—31 subjects, LS—34 subjects, TF—35 subjects.

Procedure. The subjects were instructed to read the text twice: First, to get acquainted with the experimental situation and with the text material. Second, to read the text and to underline words and/or sentences they thought were important based on their instruction. The time-limit for reading and underlining was set to 15 minutes.

Immediately after reading the text, the subject predicted how many of the underlinings he/she would be able to recall. The prediction ratings were made on a 10 cm long scale (1–100%). At the left end of the scale *none* was written, and, at the right end of the scale *all* was written. No text was written in between, and the subjects were to mark their prediction rating anywhere on the scale. Before text recall, the subjects made ratings in response to three overall text judgement questions concerning text difficulty, familiarity, and effort requirements. The same type of scale was used as for prediction ratings. The questions were as follows: *"Rate how difficult the text was to read and comprehend"*, the scale ranged from *"Not difficult at all"* to *"Very difficult"*; *"How often do you read this type of text material"*, the scale ranged from *"Never"* to *"Very often"*; *"How demanding was it to read and study the text as you have just done"*, the scale ranged from *"Not demanding at all"* to *"Very demanding"*.

The subjects recalled as much as they could remember of the text, and it was also emphasized that they should try to recall as many underlinings as possible. There was no time limit for recall. Before ending the first experimental test session, the subjects predicted their recall performance of the underlinings after a week (the same type of scale as previously). After a week's delay, the subjects recalled the text once more. They were asked to recall as much as possible, in particular the underlinings.

Results

The data from this experiment were analysed in three ways for each section: At the overall level, and for each experimental condition, pooled over achievement level. Reading levels were only studied pooled over condition. Unless otherwise noted, all results presented are signficant beyond $p < 0.05$; t-test.

Overall text judgements. At the overall level no significant differences were found in ratings of text difficulty between the four experimental conditions. The texts were regarded as rather easy to comprehend, in that the mean ratings varied between 10 and 17 (maximum 100), $F(1,125) = 1.25$, $p = 0.29$. However, a one-factor ANOVA on ratings of familiarity, with text/instruction as between subjects factor, revealed that the LS-condition was regarded as more familiar than all the other conditions. The mean for LS was 46 compared to 30 for TS, 24 for LF, and 25 for TF, $F(3,125) = 6.46$, $p < 0.05$. A second one-factor ANOVA on ratings of effort requirements, with text/instruction as between subjects factor, revealed that the LS-condition was regarded as more effort requiring than all the other three conditions. The mean for LS was 36 compared to 25 for TS, 25 for LF and 19 for TF, $F(3,125) = 3.89$, $p < 0.05$. A third one-factor ANOVA on ratings of text difficulty, with achievement level as between subjects factor, revealed that the high achievers found the texts easier to comprehend than normal and low achievers. The mean for high achievers was 7 compared to 14 for normal achievers, and 19 for low achievers, $F(2,126) = 8.41$, $p < 0.05$. A fourth one-factor ANOVA on ratings of effort requirements, with achievement level as between subjects factor, revealed that the high achievers found the texts less effort requiring than normal and low achievers. The mean for high achievers was 18 compared to 27 for normal achievers, and 33 for low achievers, $F(2,126) = 5.41$, $p < 0.05$. No difference was found between normal and low achievers' ratings of text difficulty or effort requirements.

Subjects' ratings of text difficulty, familiarity and effort were correlated with each other. Pearsons' correlation coefficient showed no significant relation between text difficulty and familiarity $r = -0.14$. A significant relation was found between text difficulty and effort, $r = 0.41$, $p < 0.05$. Familiarity and effort did not correlate, $r = 0.03$.

Prediction rating. The prediction ratings made both in the immediate and the delayed test sessions did not differ between the experimental conditions, neither for high nor for normal achievers (immediate: 46–50%; delayed: 28–33%). Two separate one-factor ANOVAs on immediate and delayed prediction ratings, with achievement level as between subjects factor, revealed that the low achievers' prediction ratings were significantly lower than the upper reading levels, for immediate recall, 39% as opposed to 49% for normal achievers, and 53% for high achievers, $F(2,126) = 6.41$, $p < 0.05$, as well as low achiever's prediction ratings of delayed recall, 23% as opposed to 32% for normal achievers, and 37% for high achievers, $F(2,126) = 5.29$, $p < 0.05$.

Recall. The subjects' recall scores were constituted by the proportion recalled underlinings out of the total number of underlinings made. Lenient scoring-criteria were used rather than demands on verbatim recall (see Recall, Experiment 1). Three categories were used for classification and scoring of the underlinings: one word, part of a sentence and a whole sentence. Recall performance was calculated by dividing recalled underlinings with the total amount of underlinings.

As can be seen in Table 3, the underlinings differ in one main respect. A one-factor ANOVA on underlinings, with text/instruction as between subjects variable, showed that the subjects who read the school-book text, regardless of instruction (LS and TS), made significantly more underlinings than did the subjects who read the fairy-tale (LF and TF) $F(3,125) = 9.75$, $p < 0.05$. The school-book text consists of more facts and it should therefore be easier to "pick out" single elements that are important. A fairy-tale, on the other hand can not so easily be broken down into single elements, but consists of rather large units of importance. Immediate recall performance was found to be almost the same for each experimental condition, 50–58%, $F(3,125) = 0.53$, $p = 0.63$. The delay reduced the recall performance significantly to 32–40% for all experimental conditions, $F(3,125) = 0.79$, $p = 0.50$. Two separate one-factor ANOVA's on immediate and delayed recall, with achievement level as between subjects factor, revealed that the high achievers recalled significantly more than normal and low achievers, 63% as opposed to 53% for normal achievers and 41% for low achievers, $F(2,126) = 12,72$, $p < 0.05$. The same pattern was found for high

Table 3. *The mean number of underlinings and the mean number of recalled underlinings for each experimental condition in Experiment 2. Recall performance in % (the ratio between recalled underlinings out of total number of underlinings made during reading), for the immediate and delayed test session. (OW = one word; PS = part of sentence; WS = whole sentence)*

	Underlinings				% immediate recall				% delayed recall			
Conditions	OW	PS	WS	Total	OW	PS	WS	Total	OW	PS	WS	Total
LS	18.7	18.4	1.8	38.9	50	54	39	58	38	32	17	40
TS	15.6	25.0	5.2	45.8	59	46	42	53	40	29	21	31
LF	3.7	17.6	1.7	23.0	62	54	29	51	49	40	12	36
TF	1.7	17.0	4.8	23.5	71	56	37	53	47	38	21	34

achievers' delayed recall, 47% as opposed to 33% for normal achievers, and 24% for low achievers, $F(2,126) = 16.13$, $p < 0.05$. Normal achievers immediate and delayed recall performance was in turn significantly better than the low achievers'. To study interactions between variables, a 3-factor ANOVA was computed, with instruction and type of text as between subjects factors, and time of test as a within subjects variable. This analysis only reconfirmed the main effect of time of test, $F(1,125) = 194.73$, $p < 0.05$, $MSe = 2.11$. No interactions for the analysis, or interactions with the achievement level variable in two subsequent ANOVAs, with data either pooled over type of text or instructions, were found to be significant.

Prediction accuracy. Prediction accuracy was calculated by means of Pearsons' correlation coefficient between the subjects' mean prediction ratings and mean recall. At the overall level, prediction ratings were found to be rather accurate for immediate recall ($r = 0.44$). The delay reduced the prediction accuracy somewhat, but was still found to be significant ($r = 0.27$). The data for each of the experimental conditions are presented in Table 4.

High prediction accuracy was found for the LS-condition both for immediate and delayed recall, whereas no prediction accuracy was found for the TF-condition. The remaining conditions, TS and LF, demonstrated high prediction accuracy for immediate recall, but no prediction accuracy was found for the delayed recall. In addition, high achievers could not accurately predict their immediate or delayed recall performance. Normal achievers accurately predicted their immediate recall performance ($r = 0.50$), but not their delayed recall performance. The low achievers could not predict their immediate recall performance, but instead their delayed recall performance ($r = 0.40$).

Table 4. *Correlation coefficients between mean prediction ratings and mean recall for each experimental condition in Experiment 2*

	Prediction accuracy	
Condition	Immediate	Delayed
LS	0.54**	0.54**
TS	0.64***	0.31
LF	0.55**	0.24
TF	0.11	0.13

$^{**}p < 0.01$
$^{***}p < 0.001$

DISCUSSION

The attempt to manipulate familiarity with the reading-task turned out to be successful in the sense that prediction accuracy was substantial for the LS-condition, both in the immediate and delayed test session. The overall text judgement questions independently support this result in that the subjects in the LS-condition made the highest ratings of familiarity. In the immediate test session, the TS- and LF-conditions also demonstrated high prediction accuracy, but not in the delayed test session. Both these conditions share one feature each with the school situation: In the TS-condition a school-book text is used, and in the LF-condition the learning instruction is advocated. However, the TF-condition had no apparent connection with the familiar school situation, and therefore, no prediction accuracy was found. As Table 3 showed, the amount of underlinings differed between the experimental conditions, and to assure that prediction accuracy was not due to a certain underlining strategy, rather than to the accuracy of the prediction rating *per se*, partial correlation coefficients were calculated for the subjective ratings, partialling out amount of recall and total amount of recall and total amount of underlinings separately. These analyses did not reveal any inflation in the prediction accuracy correlations.

The fact that students have difficulties in solving or dealing with unfamiliar problems and situations have been shown for other task-domains. For example, Säljö and Wyndhamn (1990) asked 12 and 13 year old pupils in a classroom situation to solve the everyday problem of finding the correct postage rate. The subjects were given a letter-scale and a postage table. Although the task seemed easy, it took the subjects a long time to solve the problem mainly because they approached it as they do any normal school-task.

In Experiment 1 subjects tended to overestimate their recall performance. In Experiment 2 the opposite pattern was found. Overall, 40% of the prediction ratings for immediate recalls were underestimations, and 19% overestimations. For delayed recall 40% were underestimations, and 25% overestimations. If the experimental conditions are studied separately, most underestimations were made by the subjects in the LS-condition: 55% for immediate, 48% for delayed recall (TS: 32%/35%, LF: 41%/32% and TF: 34%/43%). In the LS-condition 14% and 10% were overestimations for immediate and delayed recall respectively (TS: 13%/26%, LF: 18%/32% and TF: 25%/34%). Among high achievers 50% underestimated their immediate and 48% their delayed recall performance. Only 11% of the high achievers overestimated their immediate recall performance, but 23% of their delayed. Normal achievers followed the pattern of the high achievers in that overestimations increased with delay: 20% overestimated immediate and 29% delayed recall performance, 35% underestimated immediate and 43% delayed recall performance. The low achievers' under- and overestimations were more equally distributed. For immediate recall performance 36% underestimated and 25% overestimated. For delayed recall performance 25% underestimated and 22% overestimated. Low achievers seem to be more aware that delay will reduce their recall performance. Thus, high and normal achievers seem less aware that recall performance decline with time.

Experiment 1 and 2 revealed no systematic differences in prediction accuracy for the achievement levels. The high achievers' recall performance was better, and so were their verbal and memory abilities, but they did not seem to be aware of this fact. Maki and Berry (1984) showed that high achieving students predict their test results better than do low achieving students. However, their subjects studied what was subsequently tested during a course period. Time was therefore available for an analytic reading behavior, compared to the rather short reading-times in the present study. Thus, being efficient "metacognizers" (Sinkavich, 1988) might require a certain amount of time. Low achievers, as opposed to high achievers, would not benefit from more study time since they do not show the same strategic reading behavior (Otto, 1985). However, as mentioned, others have found that high achievers are no better than low achievers in predicting recall performance (e.g., Maki & Swett, 1987; Pressley *et al.*, 1987). High achievers perform better (e.g., Haneggi & Perfetti, 1992), and show more incidental learning (e.g., Necka *et al.*, 1992), but does this lead to a greater awareness of their own cognitive functions? According to LaBerge and Samuels (1985), a skilled reader masters different reading subskills at an automatic level. Therefore, when a skilled reader is asked about his/her every-day reading processes it will often be from a wholistic point of view, rather than from an analysis of separate steps (e.g., letter-sound, blending). One way to make students more aware of their reading processes is to increase encoding demands (e.g., Maki *et al.*, 1990). The subjects' ratings on the overall text judgement questions showed that the high achievers found the experimental demands less effortful than the other achievement groups. Therefore, they might have solved the task more automatically and fluently. Thus, the prize for being skilled could be less awareness of cognitive functions.

GENERAL CONCLUSIONS

The main purpose of the present study was to further analyse prediction accuracy of text recall. In two experiments, 9th graders predicted their recall performance of text passages. Experiment 1 could neither confirm that the best recall performance is attained with inconsistent prose passages, nor that inconsistency leads to better prediction accuracy, compared to consistent text material (Maki & Swett, 1987). On the whole, no correspondence was found between prediction accuracy of text recall and what was intended to constitute an example of ease of processing (Begg *et al.*, 1989).

However, Experiment 1 demonstrated that subjects' word knowledge, verbal inductive ability, working memory and reading time constituted significant objective predictors of recall. Verbal ability and reading time were especially crucial for low achieving students. Hence, a good working memory capacity is of minor importance if the understanding of words and language is not there. The correlation coefficient between reading time and recall performance was substantial for both high and low achieving students (high achievers $r = 0.50$; low achievers $r = 0.70$). Yet, reading time and prediction ratings were not correlated, ($r = -0.05$), which suggests that memory but not metamemory gains from more reading time.

In Experiment 2, overall prediction accuracy was found. High and long-lasting prediction accuracy was attained when the experimental conditions resembled a reading-task familiar to the subjects. The results unequivocally showed that 9th graders reached substantial prediction accuracy when they read school-book text materials with the instruction to learn. Even after a week's delay between prediction rating and the following recall test session, accuracy was high. Immediate prediction accuracy was also substantial for familiar text material (TS), and a familiar instruction (LF), but prediction accuracy was not found to be long-lasting for these experimental conditions. Experiment 2, thus revealed the encouraging result that pupils can predict their recall performance accurately, given a school-like situation. However, if the school-like situation is replaced with a less familiar situation (TF), prediction accuracy is substantially reduced. Depending on one's general approach to learning, this result may be seen as either justifying or disqualifying a given course curriculum.

In both experiments 15 year old subjects were used. They varied in achievement level, and they were asked to do the same thing, namely to predict recall performance of text. Yet, different results were attained. The procedural differences between Experiment 1 and 2 have been discussed in terms of ease, effort and familiarity. Although, ease and effort were not manipulated within the experiments, there are data that point in the direction that Experiment 1 can be viewed as a situation of easy processing, and Experiment 2 as a situation demanding increased and active processing. Four pertinent factors will be discussed.

The first factor concerns reading-time. In both experiments the subjects were instructed beforehand, that they would be asked to predict their recall performance, and also to recall. In spite of that, the subjects in Experiment 1 only used a short time, on the average 25 sec, to read each paragraph. They could have used the total time of 2 minutes. The relative ease of the texts could have caused lack of appreciation of the importance of repetition to be able to remember. Almost half of the subjects, 49%, believed that they would recall more than they actually did. In Experiment 2, reading-time was not self-paced. All the subjects were given 15 minutes to read the texts and make their underlinings. Overall, they were better able to predict their recall performance. Here, the subjects underestimated rather than overestimated their recall performance. At an overall level, 40% under- and 19% overestimated immediate recall, whereas 40% under- and 25% overestimated delayed recall.

The second factor concerns recall. The subjects in Experiment 1 read the text, paragraph by paragraph. Each passage, as presented should be recalled. In Experiment 2, the subjects personally selected their own recall key. Instead of the experimenter deciding what should be recalled, the subjects told the experimenter what was to be recalled (Greenwald & Banaji, 1989). It is argued that the underlining procedure, employed in Experiment 2, represents a condition in which more active and effort requiring reading is demanded, than just reading as in Experiment 1.

The third factor concerns the different number of prediction ratings made. In Experiment 1, the subjects made prediction ratings for each paragraphs separately (Maki & Swett, 1987). In Experiment 2, the subjects made one global prediction rating pertaining to the whole text. Pressley *et al.* (1987) argued that the subjects may have difficulties evaluating understanding and recall unless the whole text is at hand. Thus, subjects should make global estimates of recall performance.

The fourth and final factor concerns the scale being used. The scale in Experiment 2 did separate high achievers from low achievers on ratings of text difficulty, which the 7-point scale in Experiment 1 did not do. This suggested that both achievement groups in Experiment 1 found the texts rather easy to comprehend, not least since both groups on average spent little time on reading. The low achievers in Experiment 2 found the texts more difficult than high achievers, and they also rated the experimental conditions as more effortful.

The 7-point rating scale (or 5-point scale) is most typically used in this line of research (e.g., Maki & Swett, 1987; Begg *et al.*, 1992). The %-scale in Experiment 2 has not been used as often. The two different scales being used led to a somewhat confusing result. In Experiment 1 the ratio between rated text difficulty and maximum value was 40%. In Experiment 2, the same question received ratings between 10 and 17%. We argue that the effort requirements used in Experiment 2 was represented by increased text processing (i.e. underlinings) which also required more active reading. The reading-task therefore does not have to be difficult to comprehend to be effortful.

Thus, both experiments carry the following general conclusion: As long as readers deal with easy, effortless or unfamiliar tasks, awareness about cognitive functions are limited and not really needed. The more familiar the situations and the more effort you put into a task, memory awareness is enhanced as well as more explicitly demanded (Maki *et al.*, 1990).

It seems that future research in the area of prediction accuracy should further manipulate the exact relations between familiarity and effort. The present study suggested that it was the *combination* of the two that yielded the most reliable and long-lasting prediction accuracy. And, such factors may indeed be reflected in the strategic reading behavior of the individual—and should be pursued further.

The present research was supported by Grant No. 6350 from the Swedish Board of Education, awarded to the second author. Runa Patel and Ulla Britt Persson is thanked for proof-reading of the manuscript. The authors also wish to thank two anonymous reviewers for their helpful comments.

REFERENCES

Ausmeel, H. (1988). *TIPS (Text-Information Processing System): A user's guide*. Manuscript. Department of Education and Psychology, Linköping University, Sweden.

Baddeley, A., Logie, R., Nimmo-Smith, I. & Brereton, N. (1985). Components of fluent reading. *Journal of Memory and Language, 24,* 119–131.

Begg, I., Duft, S., LaLonde, P., Melnick, R. & Sanvito J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language, 28,* 610–632.

Begg, I. M., Martin, L. A. & Needham, D. R. (1992). Memory monitoring: How useful is self-knowledge about memory? *European Journal of Cognitive Psychology, 4,* 195–218.

Byrd, D. M. & Gholson, B. (1985). Reading, memory, and metacognition. *Journal of Educational Psychology, 77,* 428–436.

Cross, D. R. & Paris, S. G. (1988). Developmental and instructional analyses of children's metacognition and reading comprehension. *Journal of Educational Psychology, 80,* 131–142.

Costa, A. L. (November 1984). Mediating the metacognitive. *Educational Leadership,* 57–62.

Dermody, M. (1988). *Metacognitive strategies for development of reading comprehension for younger children.* Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education (New Orleans, LA, February 17–20, 1988).

van Dijk, T. A. & Kintsch, W. (1983). *Strategies of discourse comprehension.* London: Academic Press.

Dixon, R. A. & Hultsch, D. F. (1983). Structure and development of metamemory in adulthood. *Journal of Gerontology, 38,* 682–688.

Epstein, W., Glenberg, A. M. & Bradley, M. M. (1984). Coactivation and comprehension: Contribution of text variables to the illusion of knowing. *Memory & Cognition, 12,* 355–360.

Eysenck, M. W. & Eysenck, M. C. (1980). Effects of processing depth, distinctiveness and word frequency on retention. *British Journal of Psychology, 71,* 263–274.

Glenberg, A. M., Wilkinson, A. C. & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition, 10,* 597–602.

Glenberg, A. M. & Epstein W. (1987). Inexpert calibration of comprehension. *Memory & Cognition, 15,* 84–93.

Glenberg, A. M., Sanocki, T., Epstein, W. & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General, 116,* 119–136.

Greenwald, A. G. & Banaji, M. R. (1989). The self as a memory system: powerful but ordinary. *Journal of Personality and Social Psychology, 57,* 41–54.

Haenggi, D. & Perfetti, C. A. (1992). Individual differences in reprocessing of text. *Journal of Educational Psychology, 2,* 182–192.

Horgan, D. (1990). *Students' prediction of test grades: calibration and metacognition.* Paper presented at the Annual Meeting of the American Research Association (Boston, MA, April 16–20, 1990).

Jackson, M. D. & McLelland, J. L. (1979). Processing determinants of reading speed. *Journal of Experimental Psychology: General, 108,* 151–181.

Kahnberg, A. & Lindeberg, G. (1964). *Genom tiderna.* Lund: CWK Gleerups Förlag.

Laberge, D. & Samuels, S. J. (1985). Toward a theory of automatic information processing in reading. In Singer H. Singer & R. B. Ruddell (Eds), Theoretical models and processes of reading, third edition. Newark, DE: IRA, Inc.

Lovelace, E. A. (1984). Metamemory: Monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 756–766.

Maki, R. H. & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 663–679.

Maki, R. H. & Swett, S. (1987). Metamemory for narrative text. *Memory & Cognition, 15,* 72–83.

Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C. & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 609–616.

Maki, R. H. & Serra, M. (1992). The basis of test predictions for text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 116–126.

McDaniel, M. A., Einstein, G. O., Dunay, P. K. & Cobb, R. E. (1986). Encoding difficulty and memory: Towards a unifying theory. *Journal of Memory and Language, 25,* 645–656.

Myers II, M. & Paris, S. G. (1978). Childrens' metacognitive knowledge about reading. *Journal of Educational Psychology, 70,* 680–690.

Necka, E., Machera, M. & Miklas, E. (1992). Incidental learning, intelligence, and verbal ability. *Learning and Instruction, 2,* 141–153.

O'Brien, E. J. & Meyers, J. L. (1985). When comprehension difficulty improves memory for text. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 12–21.

Otto, W. (1985). Metacognition and reading instruction. *Journal of Reading (March), 28,* 573–575.

Paris, S. G. & Meyers II, M. (1981). Comprehension monitoring, memory, and study strategies of good and poor readers. *Journal of Reading Behavior, 13,* 5–22.

Persson, U-B. (1990). "When I read a story I make little pictures in my head". About pupils' awareness of text structure and strategy use in reading. In *Reading at the crossroads.* Proceedings of the 6th European Conference on Reading, Berlin, 1990.

Pressley, M., Snyder, B. L., Levin, J. R., Murray, H. G. & Ghatala, E. S. (1987). Perceived readiness for examination performance (PREP) produced by initial reading of text and text containing adjunct questions. *Reading Research Quarterly* Spring, 219–236.

Pullyblank, J., Bisanz, J., Scott, C. & Champion, M. A. (1985). Developmental invariance in the effects of functional self-knowledge on memory. *Child Development, 56,* 1447–1454.

Rabinowitz, J. C., Ackerman, B. P., Craik, F. I. M. & Hinchley, J. L. (1982). Aging and metamemory: the roles of relatedness and imagery. *Journal of Gerontology, 37,* 688–695.

Rönnberg, J. (1990). Cognitive and communicative function: The effects of chronological age and "handicap age". *European Journal of Cognitive Psychology, 2,* 253–273.

Rönnberg, J., Arlinger, S., Lyxell, B. & Kinnefors, C. (1989). Visual evoked potentials: Relation to adult speechreading and cognitive function. *Journal of Speech and Hearing Research, 32,* 725–735.

Sinkavich, F. J. (1988). *Multiple choice examination performance and metamemory accuracy in good and poor students.* Research report, U.S.; Pennsylvania, RIEMAY89.

Suzuki-Slakter, N. S. (1988). Elaboration and metamemory during adolescence. *Contemporary Educational Psychology, 13,* 206–220.

Säljö, R. & Wyndhamn, J. (1990). Problem-solving, academic performance and situated reasoning. A study of joint cognitive activity in the formal setting. *British Journal of Educational Psychology, 60,* 245–254.

Thompson, C. P. & Barnett, C. (1985). Review, recitation, and memory monitoring. *Journal of Educational Psychology, 77,* 533–538.

Weaver, C. A. III. (1990). Constraining factors in calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 214–222.

Westrin (undated). WIT III-manual, Psykologiförlaget, Stockholm, Sweden.

Yussen, S. R. & Levy, V. M., Jr. (1975). Developmental changes in predicting one's own span of short-term memory. *Journal of Experimental Child Psychology, 19,* 502–508.

Received 5 April 1993.

APPENDIX 1

Experiment 1: CD (Consistent/distinctive) and IcD (inconsistent/distinctive) are basically the same story with the exception of paragraph 4, which in IcD was changed into being inconsistent.

Jan's train-travel to Malmö

1. One day, on the 15th of May, Jan was going by train from Linköping to Malmö.
2. The train Jan was taking was to depart at 9.15. His parents drove him to the trian station.
3. Jan bought some sweets and a magazine in a kiosk nearby the train station.
4. *CD: Thereafter Jan boarded the Malmö-train, which left from track 1, and his trip started.*
   IcD: Thereafter Jan sat down in the car and started his travel.
5. After a while Jan heard the conductor saying that he would like to see the tickets. Jan had put his ticket in his wallet.
6. He always kept his wallet in his jacket-pocket but now it was gone.
7. Jan started to sweat. Where had he put his wallet? Had he forgotten it at the kiosk?
8. When the conductor came to Jan he told him anxiously that he could not find his wallet, in which the ticket was.
9. The conductor told him friendly that he would help Jan look for the wallet.
10. Finally, they found it in the bag Jan had recieved when he had bought his things in the kiosk. Jan was happy and went on his trip.

APPENDIX 2

Experiment 1: CNd (consistent/not distinctive) and IcNd (inconsistent/not distinctive) are basically the same story with the exception of paragraph 6, which in IcNd was changed to being inconsistent.

The Movie Visit

1. Peter and Patrik are friends. They have many interests in common.
2. Both Peter and Patrik like sports. Together they train football twice a week.
3. Another interest they have in common is going to the cinema. The films they like to see are either facts or documentaries.
4. One day when they had nothing special to do, they decided to go to the movies.
5. In the paper they could see what films were running this week.
6. *CNd: Peter and Patrik agreed on seeing a documentary.*
   IcNd: As usual Peter and Patrik agreed on seeing a sad love-story.
7. Half an hour before the film started they were at the cinema. They bought some sweets and went and sat down in the auditorium.
8. The film was a good documentary and both Peter and Patrik were satisfied when the film ended.
9. When they got out from the cinema they felt hungry and decided to buy hamburgers.
10. They bought two hamburgers each. When they had finished eating they went home, each of them to his own place.

APPENDIX 3

The fairy-tale used in Experiment 2. Reference: *Tre starka kvinnor och andra berättelser från hela världen.* Förlaget Barrikaden, Stockholm, 1980.

The Chased Hare

Once upon a time there was an old woman who lived alone on the fringe of a large, deserted moor. People in the neighbourhood spoke a lot about evil things, spirits and other terrible phenomenan which at nights roamed about on the moor. You can be convinced that the people avoided being close to the gloomy, solitary moor when it was getting darker.

Now, it so happened that the old woman had to cross the moor once a week to get to the market in town, where she sold her eggs and butter. She usually woke up early, just before dawn, to get started. One night, when she was going to the market the following day, she went to bed early. When she got up to prepare her journey, it was still dark. She had no watch, and therefore she did not know it was before midnight. She got dressed, ate, saddled her horse and hung her baskets, containing butter and eggs, on the horse. She swept an old, shabby coat around her, thereafter she and the horse started their sleepy journey across the moor.

She had not gone far when she heard a bunch of dogs barking in the starry night, and just after that, a white hare came running towards her. When it reached her, it jumped up on a rock-ledge, beside the path, as if it wanted to say: *Please, come and help me*!

The old woman laughed a little. She thought it was exciting to cheat the dogs, so she reached out, took the crouching hare and put it in one of her baskets. Then she put the cover on and rode off. The barking got closer and suddenly she saw a headless horse gallopping towards her, surrounded by a bunch of dogs. On the horse sat a dark figure with big horns on its head. The dogs' eyes were red as fire, while the tails radiated blue flames.

It was a truly terrifying sight. The horse trembled and quivered, but the woman sat straight up and waited for the horned demon. The hare was lying in the basket and she had no intention of giving it away. But it turned out that the terrifying creatures were not too smart, because the rider asked the old woman very politely if she had seen a white hare passing by, and if so, in what direction it had been running.

No, I have not, she said convincingly. *I have not seen any white hare pass me by*, which in fact was true.

The rider spurred his headless horse, urged on the dogs and gallopped away over the moor. When they were out of sight, the woman patted her trembling horse and made everything to calm it down.

To the woman's surprise the cover on one of the baskets suddenly moved. Then it was opened. However, it was not an anxious hare that turned up but a woman, all dressed in white.

The ghostlike woman spoke with a clear voice: *Madam*, she said, *I admire your courage. You have saved me from an awful spell and now it is broken. I am not a common woman — it was my destiny to be chased centuries over the moor, at nights, by evil demons, in the shape of a hare. This was to go on until I managed to get behind their tales while they were chasing me. Thanks to your courage the spell is now broken and I can return to my own people. We will never forget you. I promise that all your cows will give you plenty of milk all year round, and that the harvest in your garden will flourish like never before. But look out for the monster and his evil spirits, because he will most certainly try to hurt you if he realizes that you have been wise enough to fool him. May happiness be on your side.*

The mystical woman disappeared never to return. But everything she had promised was fulfilled. The woman sold all her butter and all her eggs in the market that morning, and happiness continued to be on her side as far as harvest and cattle were concerned.

The demon never managed to get his revenge, despite many attempts, and the ghostlike, white lady held her guardian hand over the old woman for the rest of her life.

APPENDIX 4

The school-book text (Kahnberg & Lindeberg, 1964) used in Experiment 2.

The life of Our Ancestors During the Viking Age.

The viking farm

Most Swedish people still lived along the coasts, at lakes, and river banks. There the earth was easier to cultivate. In some places the farms were gathered in small villages, in others the farms were single.

A farm consisted of several houses. These houses had low walls made of stone or thick logs, and sloping roofs, which could even reach the ground.

The largest house was the one where the master and his family lived. The roof stood on cross-beams supported by wooden pillars. In the middle of the hall the fire burned on a hearth of stones, and the smoke seeked its way out of a hole in the roof.

Around the walls were broad benches, which could be used as beds. In the middle on one side of the wall was the seat of honour, the master's seat, with pillars on both sides. On the walls hung the mens' weapons, so that they were always in reach.

Everyday life on the viking farm

The farmer and his family, the free servants and the slaves had many things to do. The men worked in the field with plough, pick, and sickle. They were chopping wood, hunting, fishing, and guarding the cattle. The women were cooking, taking care of the cattle, and weaving cloth from the farm's wool and flax.

The slaves lived in a special house. They were the property of the master as were the cattle. Some of them had been captured at war, others had been bought or were children of slaves.

The inhabitants of the samll villages and farms lived quite a lonesome life. At long intervals came tradesmen and other strangers with news from other areas. These people were treated with hospitality by the farmer.

Blood-fend

If two men were in dispute, it could happen that it was settled with weapons. If one of them were killed, his family had to take revenge by killing the perpetrator or someone of his closest relatives. If they suceeded, it was the other family's turn to take revenge. This was called blood-fend and it could lead to two families being exterminated.

It was a long time yet before there would be laws and courts that could settle disputes between the inhabitants. There are still countries in which blood-fend occurs.

The Gods of the Northerners

Despite the braveness and fearlessness of the Northerners, they believed in Gods, who were mightier than people. A lot of things happened which could not be explained other than works of the Gods.

Thunder and lightning was by our ancestors called "Tordön". One of the most prominent Gods was called Tor. He was the people's friend and helped them against giants and other evil creatures. When there was a thunderstorm, Tor was fighting with the giants. The roar came from his carriage, which was drawn by white he-goats. When he threw his hammar against the giants, there was lightning.

Another God was called Oden. He was assumed to be a one-eyed old man. He was the wisest of the Gods and to him the people brought offerings to win a battle. To his dwelling, Valhall, came all men who had been killed in a battle. Those who died of sickness or old age came to the awful land of the dead, where the Goddess Hel ruled.

Frej or Frö ruled over rain and sunshine. To him the people should sacrifice to make the seed grow in the field, and give good harvest, and to make the cattle comfortable. In spring time a picture of him was carried over the fields.

There were other Gods than these we have now been talking about. There were also Godesses married to the Gods. The Northerners believed that they looked like people, only they were bigger, stronger and did not age.

In many farms there were wooden pictures of the Gods. The people worshipped the Gods by "blota", that is sacrificing to them. Once a year the master killed a large animal, a horse or a bull, and spread the blood on the picture of the God. The meat was eaten by the people. In midwinter, when the Christians celebrated the birth of Christ, the Northerners used to sacrifice a boar to Frej to get a good harvest next year.

"Blotfester" (sacrifice parties) could also be held at holy groves or springs. There were even temples. The most famous one was in Uppsala. Every nine years people from all over the country gathered to have a large sacrifice celebration. Animals were sacrificed and even people. These offerings were then hung up in trees in a holy grove beside the temple.

Studie II

Journal of Educational Psychology
1995, Vol. 87, No. 4, 545-558

Copyright 1995 by the American Psychological Association, Inc.
0022-0663/95/$3.00

Comprehension Calibration and Recall Prediction Accuracy of Texts: Reading Skill, Reading Strategies, and Effort

Åsa Gillström and Jerker Rönnberg
Linköping University

High school students at 3 levels of verbal skill rated their own recall (prediction accuracy) and comprehension (calibration accuracy) of 3 expository texts accompanied by 3 different sets of instructions. All sets of instructions emphasized reading for understanding, and two of them also involved key words (given or personally selected), which were to be used during study. Students assessed which instructions they preferred and estimated their general verbal and memory skills. Three major results were obtained: (a) Students seemed to assess their general verbal and memory skills quite well. (b) Acceptable levels of comprehension calibration and recall prediction accuracy were found. Verbal-skill differences were found for recall prediction accuracy but not for comprehension calibration accuracy. (c) Students had study preferences—the most preferred way to study increased performance but reduced prediction accuracy.

Experiments on recall prediction and comprehension calibration accuracy were first conducted in metamemory and metacomprehension research. These experiments focused on students' ratings of their memory and comprehension abilities and compared these ratings with the students' actual performance (Maki & Berry, 1984). According to Schneider and Laurion (1993), many studies have shown that people's metacognitive skills (i.e., their ability to assess what they know or have recently learned) are not well developed. Schneider and Laurion's data showed that students accurately assessed what they knew but had problems judging what they did not know. In the present study, 16- to 19-year-old students rated their own recall and comprehension of expository texts. These subjective ratings of performance, recall predictions, and comprehension calibrations were compared with actual performance.

One question within the area of reading comprehension and recall is what effect different types of reading strategies have on performance and students' ability to rate their performance. Wade and Trathen (1989) concluded that research is inconsistent with regard to the effectiveness of teaching students optimal ways to study texts. Thus, Haenggi and Perfetti (1992) found that rereading a text, rewriting a text, and rereading notes were all equally effective in improving comprehension. Kiewra, Mayer, Christensen, Sung-Il, and Risch (1991) presented students with a videotaped lecture and found that those who took notes and those who only listened performed equally well on a recall test. Maki and Serra (1992) found that neither performance on a multiple-choice test nor recall prediction accuracy improved with practice before test taking.

However, other research has shown that increased personal involvement and effort requirements improve performance, recall prediction, and comprehension calibration accuracy. Schneider and Laurion (1993) showed that after listening to a news broadcast over the radio, students in high-involvement conditions performed better than students in low-involvement conditions. Regarding increased effort requirements, in a study of the prediction accuracy of word recall, Begg, Martin, and Needham (1992) found that students in a cued-review condition (*railroad–?*) were able to predict their performance substantially better than students in a word-pairs condition (*railroad–mother*). McDaniel (1984) found that students who read texts with deleted letters recalled more than students who read intact texts. Maki, Foley, Kajer, Thompson, and Willert (1990) showed that this type of effortful reading also improved comprehension calibration accuracy. In a similar vein, Gillström and Rönnberg (1994) instructed students to assume the role of either learners or teachers and to underline words, sentences, or both in schoolbook texts or fairy tales according to their assigned role. Recall prediction was most accurate and longest lasting (i.e., 1 week) for learner students reading schoolbook texts, a condition regarded as being more familiar and requiring more effort.

Research in the area of metacognition and verbal skill has produced conflicting results (Pressley, Snyder, Levin, Murray, & Ghatala, 1987). Gillström and Rönnberg (1994) found that although good readers had better recall performance than poor readers, poor readers had better recall prediction accuracy than good readers. Others have found the opposite relation (e.g., Garner, 1987; Maki & Berry, 1984) or that there were no differences between poor and good readers (Maki & Swett, 1987; Pressley et al., 1987).

Åsa Gillström and Jerker Rönnberg, Department of Education and Psychology, Linköping University, Linköping, Sweden.

We thank Trevor Archer, Department of Psychology, Göteborg University, and Stefan Samuelsson, Department of Education and Psychology, Linköping University, for their helpful comments and suggestions on an earlier version of this article. We also thank Karen Williams for proofreading the article.

Correspondence concerning this article should be addressed to Åsa Gillström, Department of Education and Psychology, Linköping University, S-581 83, Linköping, Sweden. Electronic mail may be sent to AsaGi@IPP.LiU.SE.

Cull and Zechmeister (1994) have suggested that methodological factors such as test material, familiarity with the task, and how performance is measured may affect results. Also, the design of the experiment—that is, whether students make multiple performance ratings per text (i.e., within-subjects) or a single global rating (i.e., between-subjects)—may play a role. According to Pressley et al. (1987), assessment of test preparedness requires global ratings of the texts rather than multiple text-segment analyses. Gillström and Rönnberg (1994) found that good readers are relatively poor at making global performance ratings and argued that this could be because good readers read in a more automatized fashion than poor readers (LaBerge & Samuels, 1985). According to Ackerman (1990), differences in general predictive abilities could be a function of attention, because the completion of a task requires more attention from a beginner than from someone skilled in the task. Thus, to complete a reading task, poor readers (like beginners) must pay more attention while reading, which results in better recall prediction accuracy (Gillström & Rönnberg, 1994).

Some studies suggest that poor and good students differ in skill and performance but not in their metacognitive abilities. Wade and Trathen (1989) found that poor students identified the important concepts in texts as well as good students but that poor students were less able to learn from the texts. McBride-Chang, Manis, Seidenberg, Custodio, and Doi (1993) found that poor and good readers scored equally well on a questionnaire assessing metacognitive awareness of reading. Thus, good students seem to have performance and learning skills that are not influenced by metacognition (Otero, Campanario, & Hopkins, 1992). In this vein, Haenggi and Perfetti (1992) concluded that proficient readers learned more from a text because they were able to relate new facts to their acquired knowledge base. Haenggi and Perfetti showed that reading instructions increased the learning performance of the less skilled readers. Cull and Zechmeister (1994) found that both poor and good readers used correct metamemory strategies in an attempt to improve learning. However, although the poor readers studied the critical items as much as or more than the good readers, their recall of these items was worse.

According to Rueda and Mehan (1986), people often use metacognitive actions in an attempt to avoid revealing incompetence. Thus, to avoid an embarrassing situation, a person must plan, monitor, evaluate, and revise his or her actions. In terms of reading, successful metacognitive strategies can result in a person passing as a reader without actually being able to read. In a study by Rueda and Mehan (1986), students with learning disabilities often completed difficult tasks with a great deal of skill. For instance, one of their students knew that he would have problems reading recipes when he took part in a cooking club. To conceal his handicap, he focused on managing two things simultaneously: his identity and the intellectual task of following the recipe without reading it. Thus, while acting as if he read the recipe, this student controlled the situation by working with others, watching others, following their lead, and imitating their actions.

The present study investigated the influence of effort and personal involvement on the metamemory and metacomprehension abilities of students at three different levels of verbal skill. Verbal skill was assessed with two tests, and students rated their general reading fluency and comprehension and their memory. Students read three expository texts accompanied by three different sets of instructions about how to read to enhance comprehension. For two sets of the instructions, the students were told to use key words; in one case the key words were given, and in the other they were selected by the students. The effects of instruction type on recall and comprehension performance were tested.

We tested the following four hypotheses: First, students at all levels of verbal skill should demonstrate equal awareness of their verbal and memory abilities and thus should be equally able to estimate these abilities (Cull & Zechmeister, 1994; McBride-Chang et al., 1993; Persson, 1994; Rueda & Mehan, 1986; Wade & Trathen, 1989). Second, there should be effects of instruction type such that the use of key words, which requires more effort and active processing, leads to increased recall and comprehension performance, increased recall prediction, and increased comprehension calibration accuracy (Maki et al., 1990; McDaniel, 1984). Higher ratings of effort would confirm the validity of the instructional manipulation. Third, further effects of instruction type should show that the use of self-selected key words results in the highest recall prediction accuracy scores of the three instructional conditions, because the self-selection of key words is presumed to require more personal involvement. Fourth, low-verbal-skill students should show higher recall prediction and comprehension calibration accuracy than high-verbal-skill students because low-verbal-skill students must be more attentive while reading, whereas high-verbal-skill students read in a highly automatized manner (Ackerman, 1990; Gillström & Rönnberg, 1994; LaBerge & Samuels, 1985).

Method

Participants

A total of 111 Swedish high school students were divided into three levels of verbal skill on the basis of their performance on two verbal tests (see *Phase 1*, below). Their mean age was 17.4 years (*SD* = 0.95). Previous research has shown that verbal test scores and teacher ratings of reading skill are highly correlated (cf. Gillström & Rönnberg, 1994; Necka, Machera, & Miklas, 1992; Wade & Trathen, 1989).

Text Materials

Three short expository texts were used in the experiment (see Appendix). The texts were taken from a standardized diagnostic test battery consisting of five separate tests, which are used in Sweden for the assessment of the degree of reading and writing difficulties experienced by students in Grades 7–9 (Psykologiförlaget, 1976). One of the tests assesses reading comprehension. This test consists of 14 similar short expository texts accompanied by 30 multiple-choice questions (2 or 3 questions for each text). Students were given 30 min to study the texts.

So that we could select three texts, 19 students read the 14 texts and answered the reading comprehension questions following each text. The experimental texts chosen resulted in about 50% correct performance on the multiple-choice questions and were equally long, approximately 135 words. This implies that the chosen texts were sufficiently demanding (Gillström & Rönnberg, 1994). The chosen texts were labeled Text 1: Banana, Text 2: Arabia, and Text 3: Indian (see Appendix).

Instructions

The three instructions were as follows:

> 1. Read to understand (READING): You will now read the text through until you feel that you have understood what it is all about. After 4 minutes you are going to answer two questions concerning your comprehension of the text, and you will also try to recall as much of the text as possible.
>
> 2. Read to understand and use the 5 given words from the text, as support (GIVEN): You will now read the text through until you feel that you have understood what it is all about. Below the text you will find 5 words selected from the text which you can use as support when, after 4 minutes, you are going to answer two questions concerning your comprehension of the text, and you will also try to recall as much of the text as possible.

Selection of words used in the GIVEN instructions was done by 11 university teachers in psychology. They were instructed to read the texts and to select 5 words that would support recall and comprehension. The most typically selected words were chosen as key words in the experiment.

> 3. Read to understand and select 5 words of your own from the text as support (SELECTED): You will now read the text through until you feel that you have understood what it is all about. As an aid, you can select 5 words from the text which you can use as support when, after 4 minutes, you are going to answer two questions concerning your comprehension of the text, and you will also try to recall as much of the text as possible.

Both reading time and time for recall were limited to 4 min and 5 min, respectively. Mazzoni and Cornoldi (1993) showed that having more time to study does not necessarily increase text recall performance.

Order of Texts and Instructions

To minimize the possibility of carryover or sequence effects and to be able to view instructional effects pooled over texts, we counterbalanced the order of texts and instructions across the students. Students were randomly assigned to one of three orders of presentation of texts and instructions. A split-plot experimental design was used, with order of presentation as a between-subjects factor and texts and instructions as within-subjects factors. The layout of the design followed a Latin square procedure (Kirk, 1968; see Table 1). Subsequent univariate Separate Order × Text analyses of variance (ANOVAs), with recall, recall prediction, recall postdiction, comprehension, and confidence ratings as dependent variables, revealed no order of presentation main effects. (Recall postdiction ratings are those ratings the students made after having recalled. That is, the students estimated how much they actually were able to recall.) Because Order × Instructions interactions were not meaningful to the issue of carryover effects, we ignored order in all analyses reported here.

General Design and Procedure

The experiment took place in the classroom during high school psychology lectures. All three orders of instructions and texts were represented within each classroom. Approximately 20 students participated at one time. The experiment took 80 min. Each student was given a booklet that contained all questions, experimental texts, instructions, space for recall, and so forth. On the front page, the students filled in personal background data. Students could voluntarily fill in their grade point average and their grades in Swedish language. After completion of the experiment, the primary experimenter was given 40 min to inform the students about the purpose of the experiment. The experiment consisted of three phases.

Phase 1. The students' verbal ability was measured by two verbal tests: one analogy test and one synonym–antonym test. As with the experimental texts, the verbal tests belong to a standardized test battery that is used in Sweden to measure study success (Westrin, 1965). For ninth graders, the average score on both of these tests, combined, was 15.1 (for analogy, 15.5 out of 27; for synonym–antonym, 14.7 out of 29), and the correlation was .71. After completing the verbal tests, the students rated their general reading and memory abilities using three overall rating items.

The analogy test measures verbal inductive ability. As a practice example, the students were presented with the word pair *driver–car* and were instructed to find an analogous pair out of the following five words: *trot, riding, horse, ride,* and *rider*. The test consists of 27 similar items, and the students were given 5.5 min to complete the test. The synonym–antonym test measures word knowledge. As a practice example, the students were presented with the following five words: *false, rare, erroneous, genuine,* and *whole*. The students were to mark the two words that are opposite in meaning. The test consists of 29 items, and the students were given 5 min to complete the test.

Table 1
Three Presentation Orders of Texts and Instructions

Presentation order	Instructions: Text title		
	First reading	Second reading	Third reading
First	READING: Banana	GIVEN: Arabia	SELECTED: Indian
Second	SELECTED: Arabia	READING: Indian	GIVEN: Banana
Third	GIVEN: Indian	SELECTED: Banana	READING: Arabia

Note. Thirty-seven students were randomly assigned to each presentation order. READING = instructions were "read to understand," GIVEN = instructions were "read to understand and use the five given words from the text as support," SELECTED = instructions were "read to understand and select five words of your own from the text as support."

The three items on reading and memory abilities (general questions) were subjective ratings of

1. Fluent reading ability: *I estimate that my reading ability is* ________.
2. Reading comprehension ability: *I estimate that my reading comprehension is* ________.
3. Memory ability: *I estimate that my memory ability is* ________.

These items were accompanied by rating scales, ranging from *very poor* (0%) to *very good* (100%); the scales were straight lines with no points or verbal indicators between the ends of the scale.[1]

Phase 2. Each text was read for 4 min, and the students noted how many readings they were able to complete. Then the students' comprehension calibration accuracy was tested. The procedure was as follows: First, on a scale ranging from *very poor* (0%) to *very good* (100%), the students rated their comprehension by completing the following item: *I estimate that my comprehension of the text is* ________. Second, as a measurement of text comprehension, the students answered the two multiple-choice questions for the text. Third, the students made confidence ratings as to the correctness of their answers on the multiple-choice questions: *Estimate how confident you are that your answer on Question 1 (or 2) is correct.* Both confidence ratings were made on a scale ranging from *very unsure* (0%) to *very confident* (100%).

The procedure to assess recall prediction accuracy was as follows: First, the students predicted their recall performance (after having read the text) on a scale ranging from *nothing* (0%) to *everything* (100%), *Estimate how much of the text you will be able to recall.* Second, for 5 min the students recalled as much as possible of the text. Third, the students postdicted their recall performance on a scale ranging from *nothing* (0%) to *everything* (100%), *Estimate how much of the text you were able to recall.*

For all three texts the procedure was the same: Assessment of comprehension calibration was followed by prediction accuracy of texts. Before the students began reading a new text with new instructions, they rated the old text or instructions in terms of comprehension difficulty and effort requirements on scales ranging from *very easy* (0%) to *very difficult* (100%) and from *no effort at all* (0%) to *very much effort* (100%), respectively: (a) *How difficult was the text to read and comprehend?* and (b) *How much effort was required to read and study the text?*.

Phase 3. In the final and more general phase, the students were asked to assess the three instruction types by responding to four questions: (a) Which instructions were the easiest to use? (b) Which instructions facilitated comprehension? (c) Which instructions facilitated recall? (d) Future processing choice, that is, which instructions would you choose again if you were to read a new text? The students were instructed to mark only one of the three instructions, and they were asked to explain their reply to Question 4.

Results

The data were analyzed separately for each phase. In most cases, these analyses were reported at the overall level and by verbal-skill level. The division into three levels of verbal skill was based on the mean number of correct answers on the synonym–antonym and analogy tests (maximum score = 28.0). Twenty students were classified as high-verbal-skill students (scoring between 21.0 and 26.0), 72 were classified as medium-verbal-skill students (scoring between 12.5 and 20.0), and 19 were classified as low-verbal-skill students (scoring between 4.5 and 12.0). The groups were chosen so that the two extreme verbal-skill levels could be studied more carefully. Data from Phase 3 were used to regroup students for post hoc analyses. These new groupings were based on students' responses to the questions evaluating readability, recallability, and instruction preference. In addition, qualitative analyses were used to elucidate and exemplify the students' evaluations.

Because the number of medium-verbal-skill students was much greater, data were first analyzed for the high- and low-verbal-skill students only. The main effects and interactions attained for these extreme skill groups did not change when the medium-verbal-skill students were included in the analyses. Therefore, all students were included in the analyses, allowing us to study how strategies, skills, and effort requirements affect students at all three skill levels. All differences among the verbal-skill groups were tested with two-tailed t tests. Calculations of correlations were based on Pearson product–moment correlations.

Phase 1

Objective measures and overall judgments. The scores on the analogy test and the synonym–antonym test were significantly correlated, $r = .69$. Thus, students with good word knowledge also showed good verbal inductive ability, and vice versa. For all students, the mean score on the analogy test was 16.5 (out of 27), and the mean score on the synonym–antonym test was 15.9 (out of 29). The mean test scores and the correlation coefficient between the tests closely corresponded to the standardized results in the WIT III manual from which these tests were taken (Westrin, 1965).

A two-factor ANOVA (based on a simultaneous regression least squares solution) on the general questions of fluency, comprehension, and memory, with verbal skill as a between-subjects factor and general questions as a within-subjects factor, revealed a verbal-skill-level main effect, $F(2, 108) = 17.02$, $p < .05$, $MSE = 544.87$, as well as a general questions main effect, $F(2, 216) = 36.49$, $p < .05$, $MSE = 204.02$. A significant interaction was detected, $F(4, 216) = 2.62$, $p < .05$, $MSE = 204.02$. As may be seen in Table 2, the primary source of this interaction is that the difference between high-verbal-skill students and either medium- or low-verbal-skill students was greater on the comprehension questions than on the fluency questions, with interaction comparisons (Marascuilo & Levin, 1970) yielding $ts = 3.05$ and 2.36 for medium- and low-verbal-skill students, respectively, both $ps < .025$.

As Table 3 shows, the verbal tests correlated moderately ($rs = .40$ and .50) with recall performance, which confirms previous data on objective prediction accuracy (Gillström & Rönnberg, 1994). Thus, those students with better word

---

[1] Throughout the experiment the same type of rating scale was used: The students could mark their ratings anywhere on a 10-cm scale. The students were told to interpret the scale in terms of percentages.

Table 2
Students' Ratings of Fluency of Reading, Reading Comprehension, and Memory Abilities

	General questions					
	Fluency		Comprehension		Memory	
Rating	*M*	*SD*	*M*	*SD*	*M*	*SD*
Overall (n = 111)	72.92	18.71	63.40	19.52	51.72	19.81
By verbal skill						
High (n = 20)	77.45	14.39	80.65	15.34	61.70	18.08
Medium (n = 72)	75.39	17.22	63.01	16.62	52.15	18.38
Low (n = 19)	58.79	22.22	46.74	19.15	39.58	21.28

knowledge and better verbal inductive ability also showed better recall performance, and vice versa. In addition, the subjective ratings of reading and memory abilities were correlated with both the verbal test results and the students' recall performance. To assess the actual accuracy of these relations, we calculated the absolute differences between the subjective ratings of verbal and memory abilities and the actual verbal test performance and recall. Ratings within the range of 20% of actual performance were regarded as accurate. With few exceptions, 50% to 72% of the subjective ratings of reading and memory abilities fell within this range. As an example, 68% of the ratings of memory ability accurately matched recall performance across instructions. Seventy-two percent and 71% of the ratings of reading comprehension accurately matched performance on the synonym–antonym test and the analogy test, respectively.

Summary of Phase 1 results. The Phase 1 data confirmed the first hypothesis in that all students were comparably accurate at estimating their own verbal and memory abilities (cf. Cull & Zechmeister, 1994; McBride-Chang et al., 1993; Wade & Trathen, 1989). According to Rueda and Mehan (1986) this result could reflect students' social awareness of their position in the school system (cf. Persson, 1994).

Phase 2

Correct answers. Actual text comprehension was measured by the number of correct answers on the two multiple-choice questions following each instruction and text. The mean score for each student was calculated. A two-factor ANOVA on mean number of correct answers, with verbal skill as a between-subjects factor and instructions as a within-subjects factor, revealed a significant verbal-skill-level main effect, $F(2, 108) = 26.39$, $p < .05$, $MSE = 0.11$, but no instructions main effect, $F(2, 216) < 1$. No interaction was found, $F(4, 216) < 1$ (see Table 4). Regardless of instructions, the high-verbal-skill students performed better (.86) than the medium- (.67) and the low-verbal-skill students (.42), $t(90) = 4.00$, $p < .05$, and $t(37) = 8.39$, $p < .05$, respectively. The medium-verbal-skill students performed better than the low-verbal-skill students, $t(89) = 4.81$, $p < .05$. An analysis conducted on the arcsine-transformed data yielded parallel results.

Comprehension calibration ratings. A two-factor ANOVA on comprehension calibrations, with verbal skill as a between-subjects factor and instructions as a within-subjects factor, revealed a significant verbal-skill-level main effect, $F(2, 108) = 23.80$, $p < .05$, $MSE = 581.38$, as well as a significant instructions main effect, $F(2, 216) = 6.68$, $p < .05$, $MSE = 245.90$ (Table 4). No interaction effect was found, $F(4, 216) < 1$. Regardless of instructions, mean ratings were significantly higher for the high-verbal-skill students (76.55) as compared with the medium- (64.31) and the low-verbal-skill students (46.05), $t(90) = 3.54$, $p < .05$, and $t(37) = 7.65$, $p < .05$, respectively. In addition, the medium-verbal-skill students made higher ratings than did the low-verbal-skill students, $t(89) = 4.81$, $p < .05$. Stu-

Table 3
Correlations Among Students' Ratings of Fluency of Reading (FR), Reading Comprehension (RC), Memory Ability (ME), Recall Performance for the Three Instructions, and the Verbal Test Results

Measure	1	2	3	4	5	6	7	8
1. FR	—							
2. RC	.53**	—						
3. ME	.29**	.50**	—					
4. Recall READING	.27**	.32**	.23*	—				
5. Recall GIVEN	.40**	.44**	.28**	.55**	—			
6. Recall SELECTED	.35**	.36**	.32**	.44**	.72**	—		
7. Analogy	.31**	.50**	.35**	.46**	.48**	.43**	—	
8. Synonym–antonym	.40**	.56**	.30**	.56**	.56**	.54**	.69**	—

Note. $n = 111$. READING = instructions were "read to understand," GIVEN = instructions were "read to understand and use the five given words from the text as support," SELECTED = instructions were "read to understand and select five words of your own from the text as support."
* $p < .05$. ** $p < .01$.

Table 4
Mean Proportion of Correct Answers, Mean Comprehension Calibrations, and Reliability of Comprehension Calibrations for the Three Instructions

	Instructions					
	READING		GIVEN		SELECTED	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*
	Proportion of correct answers					
Overall ($n = 111$)	.65	.36	.65	.36	.68	.27
By verbal skill						
High ($n = 20$)	.82	.24	.87	.22	.87	.22
Medium ($n = 72$)	.67	.36	.67	.35	.67	.34
Low ($n = 19$)	.39	.36	.37	.37	.50	.33
	Comprehension calibration %					
Overall ($n = 111$)	67.34	20.80	63.72	21.64	59.11	20.26
By verbal skill						
High ($n = 20$)	81.90	9.52	79.65	14.52	68.10	16.14
Medium ($n = 72$)	67.99	20.75	64.12	20.11	60.82	19.21
Low ($n = 19$)	49.58	16.76	45.42	20.18	43.16	20.25
	Reliability of comprehension calibrations					
Overall ($n = 111$)	.29**		.47**		.27*	
By verbal skill						
High ($n = 20$)	.15		−.19		.06	
Medium ($n = 72$)	.12		.47**		.23*	
Low ($n = 19$)	.34		.13		.01	

Note. READING = instructions were "read to understand," GIVEN = instructions were "read to understand and use the five given words from the text as support," SELECTED = instructions were "read to understand and select five words of your own from the text as support."
* $p < .05$. ** $p < .01$.

dents' mean ratings were higher with both the READING (67.34) and the GIVEN (63.72) instructions than they were with the SELECTED (59.11) instructions, $t(110) = 4.28$, $p < .05$, and $t(110) = 2.24$, $p < .05$, respectively.

Confidence ratings. A two-factor ANOVA on the confidence ratings revealed a verbal-skill-level main effect, $F(2, 108) = 22.61$, $p < .05$, $MSE = 549.86$, but no instructions main effect, $F(2, 216) < 1$. A significant interaction was detected, $F(4, 216) = 2.45$, $p = .05$, $MSE = 212.30$. Across instructions low-verbal-skill students provided significantly lower confidence ratings (47.53) than either high-verbal-skill students (74.64), $t(37) = 6.82$, $p < .05$, or medium-verbal-skill students (68.12), $t(89) = 5.65$, $p < .05$. The medium- and the high-verbal-skill students' ratings did not differ, $t(90) = 1.93$, $p > .05$. Subsequent Scheffé testing of the two-way means did not detect any substantively meaningful contrasts.

Calibration accuracy of comprehension. Overall, significant relations between calibrated comprehension and number of correctly answered questions were found across instructions. Significant relations were found for the medium-verbal-skill students with the GIVEN and SELECTED instructions (see Table 4).

To assess the actual accuracy of these ratings, we calculated the difference between the proportions of calibrated comprehension (C) and correctly answered questions (A) for all students, with smaller absolute differences (C − A) corresponding to more accurate ratings. An acceptable range of 20%, out of the total 100%, correct answers was defined as accurate comprehension calibration. Across verbal skill and as a function of instruction type, the percentage of students who attained an acceptable level of comprehension calibration accuracy was as follows: 36% (READING), 48% (GIVEN), and 40% (SELECTED). Across instructions and as a function of verbal-skill level, the comparable percentages were as follows: 35% (low), 44% (medium), and 38% (high).

Recall performance. Each of the three experimental texts was divided into 33 propositions. The mean number of recalled propositions was calculated for each student. As described by Noice (1993), the recall protocols were scored in a deviation from verbatim fashion; correct recall could include any of the following nonessential deviations from true verbatim: adding words, switching words or idea units, substituting words, adding conjunctions such as *and* and *but*, accepting any form of verb, and substituting singular form for plural form (for further details see Noice, 1993).

A two-way ANOVA on recall performance revealed a verbal-skill-level main effect, $F(2, 108) = 22.13$, $p < .05$, $MSE = 498.45$, but instructions had no apparent effect on recall performance, $F(2, 216) = 1.61$, $p > .05$, $MSE = 140.05$. No interaction was found, $F(4, 216) < 1$ (Table 5). Across instructions, the high-verbal-skill students recalled more (53.99%) than did the medium- (45.06%) and low-verbal-skill students (27.27%), $t(89) = 2.81$, $p < .05$, and $t(37) = 6.95$, $p < .05$, respectively. In turn, the medium-verbal-skill students recalled more than the low-verbal-skill students, $t(88) = 5.05$, $p < .05$.

Table 5
Percentage of Recalled Propositions, Mean Recall Predictions, and Reliability of Recall Predictions for the Three Instructions

	Instructions					
	READING		GIVEN		SELECTED	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*
			Recalled propositions %			
Overall ($n = 111$)	43.03	18.33	45.21	18.06	42.24	17.53
By verbal skill						
High ($n = 20$)	51.52	11.72	56.06	13.02	54.39	14.57
Medium ($n = 72$)	45.58	16.82	46.42	16.67	42.85	16.32
Low ($n = 19$)	24.39	17.86	29.18	17.62	28.24	15.12
			Recall predictions %			
Overall ($n = 111$)	50.69	20.09	50.04	17.35	48.78	17.38
By verbal skill						
High ($n = 20$)	58.70	17.45	63.50	15.54	58.70	14.99
Medium ($n = 72$)	51.96	19.50	49.15	15.51	47.89	16.68
Low ($n = 19$)	37.47	19.59	39.21	17.56	35.89	15.12
			Reliability of recall predictions			
Overall ($n = 111$)	.31**		.36**		.38**	
By verbal skill						
High ($n = 20$)	.04		−.11		.15	
Medium ($n = 72$)	.12		.20		.28*	
Low ($n = 19$)	.53*		.54*		.15	

Note. READING = instructions were "read to understand," GIVEN = instructions were "read to understand and use the five given words from the text as support," SELECTED = instructions were "read to understand and select five words of your own from the text as support."
* $p < .05$. ** $p < .01$.

The same two-factor ANOVA on number of readings, revealed a verbal-skill-level main effect, $F(2, 108) = 5.75$, $p < .05$, $MSE = 4.39$, and an instructions main effect, $F(2, 216) = 24.29$, $p < .05$, $MSE = 1.25$. No interaction effect was found, $F(4, 216) = 1.87$, $p > .05$, $MSE = 1.25$. Across instructions, the high-verbal-skill students read the texts more times (5.12) than did the medium- (4.23) and low-verbal-skill students (3.89), $t(89) = 2.82$, $p < .05$, and $t(37) = 2.84$, $p < .05$, respectively. No difference was found between the medium- and the low-verbal-skill students, $t(88) = 1.19$, $p > .05$. Across verbal skill, students read the texts almost an equal number of times with the READING (4.83) and GIVEN instructions (4.53), $t(110) = 1.72$, $p > .05$. With the SELECTED instructions, students read 3.66 times, which was fewer than with the READING, $t(110) = 6.72$, $p < .05$, and the GIVEN instructions, $t(110) = 5.60$, $p < .05$. The number of readings did not affect overall recall performance. Correlations between recall and number of readings were computed for each of the instructions. Only low and nonsignificant correlations were obtained; for READING $r = -.04$, for GIVEN $r = .12$, and for SELECTED $r = .19$ (cf. Mazzoni & Cornoldi, 1993).

Recall predictions. A verbal-skill main effect was found, $F(2, 108) = 14.21$, $p < .05$, $MSE = 533.95$, but instructions had no overall effect on the recall predictions, $F(2, 216) = 1.11$, $p > .05$, $MSE = 173.05$. No interaction effect was found, $F(4, 216) < 1$ (Table 5). Across instructions, mean recall predictions for the high-verbal-skill students (40.73) were higher than they were for the medium- (33.70), $t(90) = 3.11$, $p < .05$, and the low-verbal-skill students (25.56), $t(37) = 4.48$, $p < .05$. Mean recall predictions for the medium- and the low-verbal-skill students also differed, $t(89) = 3.40$, $p < .05$.

A significant main effect of verbal skill was also found for the recall postdiction data, $F(2, 108) = 16.77$, $p < .05$, $MSE = 620.70$, but not for instructions, $F(2, 216) = 2.47$, $p > .05$, $MSE = 255.35$. No interaction effect was found, $F(4, 216) < 1$. Across instructions, mean postdictions for the high-verbal-skill students (65.73) were higher than they were for the medium- (53.03), $t(90) = 3.48$, $p < .05$, and the low-verbal-skill students (39.05), $t(37) = 6.39$, $p < .05$. The mean postdictions for the medium- and the low-verbal-skill students also differed, $t(89) = 3.65$, $p < .05$.

Recall prediction accuracy. As hypothesized, the reliability of the overall correlations between recall predictions and recall performance was significant for all instructions, but it was of modest magnitude (see Table 5). No relation between predicted and actual recall was found for the high-verbal-skill students, but a relation was found for the low-verbal-skill students in two conditions (i.e., READING and GIVEN). A significant relation was found for the medium-verbal-skill students with the SELECTED instructions.

To assess the actual accuracy of these predictions, we calculated the difference between the proportions of predicted (P) and actual recall (A) for all students, with smaller absolute differences (P − A) corresponding to more accurate recall predictions. An acceptable range of 20%, out of the total 100%, over or under actual recall, was defined as accurate recall prediction. Across verbal skill, 60% to 64% of the students attained an acceptable level of recall prediction accuracy, and fewer than 7% of the students exceeded 40% recall prediction inaccuracy. Across instructions and as a function of verbal skill, the percentages of students with acceptable accuracy were as follows: 54% (low), 64% (medium), and 65% (high). Few students exceeded 40% recall prediction inaccuracy.

Recall postdictions were also correlated with recall performance. At the overall level and as a function of instruction type, correlations between recall postdictions and recall performance were as follows: .43 (READING), .47 (GIVEN), and .54 (SELECTED). Across instructions and as a function of verbal skill, correlations between recall postdictions and recall performance were as follows: between .56 and .70, $p < .01$ (low), between .25 and .51, $p < .05$ (medium), and between −.04 and .25, $p > .05$ (high). To study the actual accuracy of the postdictions, we calculated differences in the proportions (P − A). Overall, 60% to 70% of the students made accurate postdictions (i.e., fell within the range of 20% over or under actual recall). Very few exceeded 40% inaccuracy. Across instructions, 74% of the low-verbal-skill, 63% of the medium-verbal-skill, and 58% of the high-verbal-skill students made accurate postdictions. Again, very few exceeded 40% inaccuracy.

Ratings of effort and text difficulty. A two-factor ANOVA on effort requirements, with verbal skill and instructions as variables, revealed no verbal-skill main effect $F(2, 108) = 1.99, p > .05, MSE = 704.32$, but did reveal an instructions main effect, $F(2, 216) = 3.37, p < .05, MSE = 253.87$. No interaction effect, $F(4, 216) = 1.78, p > .05, MSE = 253.87$, was found. Overall mean effort ratings with the GIVEN instructions were 44.49, which was less than with the READING (48.71), $t(110) = 2.11, p < .05$ and SELECTED (50.97) instructions, $t(110) = 3.32, p < .05$. Overall mean effort ratings between READING and SELECTED did not differ, $t(110) = 1.13, p > .05$.

The same ANOVA on text difficulty ratings revealed a significant verbal-skill-level main effect, $F(2, 108) = 14.74, p < .05, MSE = 714.87$, and an instructions main effect, $F(2, 216) = 4.92, p < .05, MSE = 271.43$. No interaction effect was found, $F(4, 216) < 1$. For high-verbal-skill students, the mean ratings across instructions was 23.28, which was less than for the medium-verbal-skill students (35.01), $t(90) = 2.99, p < .05$, and for the low-verbal-skill students (50.05), $t(37) = 6.34, p < .05$. The medium-verbal-skill students made lower ratings than did the low-verbal-skill students, $t(89) = 3.61, p < .05$. Overall means for the instructions were higher for the READING (32.52) and GIVEN (34.66) instructions as compared with the SELECTED (39.31) instructions, $t(110) = 3.12, p < .05$, and $t(110) = 2.29, p < .05$.

Summary of Phase 2. Performance on the comprehension questions did not differ as a function of instructions but did differ as a function of verbal ability. Comprehension calibration ratings were affected by instructions, with SELECTED yielding the lowest mean ratings. Overall, significant correlations between calibrated comprehension and comprehension performance were found. Between 36% and 48% of the students made accurate calibrations of comprehension.

Overall recall performance did not differ as a function of instructions, but verbal skill was again an important factor. Significant correlations between recall predictions and performance were found overall and for the lower verbal-skill groups. Between 60% and 64% of the students made accurate ratings of their performance. Active processing (i.e., GIVEN and SELECTED) did not increase recall prediction accuracy. As intended, the subjective ratings of effort requirements revealed that the experimental situation was demanding (Gillström & Rönnberg, 1994; Maki et al., 1990).

Phase 3

Students' assessment of the instructions. The students answered four questions that assessed different aspects of the instructions: (a) which instructions were easiest to use (EASE), (b) which instructions facilitated comprehension (COMPREHENSION), (c) which instructions allowed the highest recall (RECALL), and (d) future processing choice, that is, which instructions students would choose if they were to read a new text (CHOICE). Answers to the fourth question required that the students motivate their choice. Table 6 shows that the distribution of students' preferences were almost equally divided among the three instructions, with the exception of COMPREHENSION. More than half of the students (54%) thought the READING instruction

Table 6
Percentage of Students' Preferences of Instructions That Were the Easiest to Use (EASE), Facilitated Comprehension (COMPREHENSION), Facilitated Recall (RECALL), and Influenced Future Processing Choice (CHOICE)

Measure and instruction	EASE	COMPREHENSION	RECALL	CHOICE
Overall				
READING	36	54	31	33
GIVEN	32	19	31	27
SELECTED	32	27	38	40
By verbal skill				
High				
READING	40	70	20	35
GIVEN	35	10	40	30
SELECTED	25	20	40	35
Medium				
READING	32	51	30	28
GIVEN	32	21	28	29
SELECTED	36	28	42	43
Low				
READING	48	53	42	53
GIVEN	26	16	32	16
SELECTED	26	31	26	31

Note. READING = instructions were "read to understand," GIVEN = instructions were "read to understand and use the five given words from the text as support," SELECTED = instructions were "read to understand and select five words of your own from the text as support."

resulted in better comprehension. However, we found that comprehension performance was more equally divided among the three instructions. Thus, the subjective evaluation of COMPREHENSION was excluded from further analyses.

To assess the validity of the other three questions, we carried out post hoc analyses that were based on the subjective responses. Thus, regardless of verbal skill, those students who thought that the READING instructions best facilitated recall performance formed one group, and their performance and ratings with that instruction were compared with their performance and ratings made with the other two instructions that they used. If students' recall performance was best for the instructions they thought best facilitated recall, then their assessment was considered accurate. The results of Questions 3 (RECALL) and 4 (CHOICE) are presented in Tables 7 and 8 (Question 1 shows the same pattern as these questions). In addition to recall predictions and actual recall performance, we included effort ratings in the analyses.

Table 7 shows that, in most cases, the students made higher recall predictions, showed lower effort ratings, and had better recall with the preferred instructions as compared with the other two sets of instructions. This pattern was found for RECALL as well as for CHOICE and EASE. These types of ratings apparently validate each other.

The correlations between predicted and actual recall for the post hoc groups are displayed in Table 8. In almost all cases, higher and significant relations were found for the nonpreferred, more effort-demanding instructions. We examined the difference in proportions (P − A) for the preferred and nonpreferred instructions to assess the accuracy of these relations. Overall, most students' recall predictions fell within the range of 20% for the nonpreferred instructions. Fewer students' recall predictions did so for the personally best instructions. For example, 41% of those students who preferred the READING instructions for RECALL made recall predictions within 20% with those instructions, whereas between 67% and 71% of the same group of students made recall predictions within 20% with the other two sets of instructions.

Qualitative analysis of the students' motivations for their

Table 7

Post Hoc Groups' Mean Recall Predictions, Effort Ratings, and Recall Performance for Question 3 (Facilitated Recall) and Question 4 (Future Processing Choice) and for the Other Two Instructions That the Students Used

		Post hoc group					
		READING		GIVEN		SELECTED	
Question	Instructions	*M*	*SD*	*M*	*SD*	*M*	*SD*
Facilitated recall							
Prediction %	READING	**55.21**$_{a}$	23.14	49.38	16.14	48.16	20.20
Prediction %	GIVEN	45.29	14.38	**55.18**$_{c}$	17.03	49.72	18.94
Prediction %	SELECTED	46.56	18.21	42.29	13.21	**53.09**$_{a}$	18.40
Effort %	READING	**39.35**$_{a,b}$	20.27	49.79	19.89	55.26	20.38
Effort %	GIVEN	49.09	18.84	**36.32**$_{b,c}$	20.09	47.33$_{b}$	20.71
Effort %	SELECTED	54.00	18.58	51.47	18.30	**48.19**$_{a}$	20.28
Recall %	READING	**46.97**$_{a}$	6.88	41.18	5.37	41.36	5.83
Recall %	GIVEN	43.64	6.25	**46.24**$_{c}$	5.98	45.61	5.83
Recall %	SELECTED	38.15	4.77	39.03	6.19	**48.48**$_{a}$	5.74
Future processing choice							
Prediction %	READING	**54.19**	22.46	50.03$_{a}$	16.44	48.20	20.27
Prediction %	GIVEN	50.30	14.67	**51.07**$_{c}$	18.05	49.11	19.20
Prediction %	SELECTED	48.19	18.16	42.87	13.75	**51.48**	17.86
Effort %	READING	**39.14**$_{a,b}$	18.58	53.84	17.81	53.29	21.76
Effort %	GIVEN	48.32	18.63	**38.03**$_{b,c}$	21.85	45.68$_{b}$	20.57
Effort %	SELECTED	55.54	17.55	53.70	17.31	**45.27**$_{a}$	20.55
Recall %	READING	**46.68**$_{a}$	5.89	42.12	6.65	40.56	5.71
Recall %	GIVEN	44.55	5.72	**42.32**	6.00	47.73	6.15
Recall %	SELECTED	39.72	4.64	36.36	6.30	**48.83**$_{a}$	5.73

Note. READING = instructions were "read to understand," GIVEN = instructions were "read to understand and use the five given words from the text as support," SELECTED = instructions were "read to understand and select five words of your own from the text as support." For facilitated recall, READING $n = 34$, GIVEN $n = 34$, and SELECTED $n = 43$. For future processing choice, READING $n = 37$, GIVEN $n = 30$, and SELECTED $n = 44$. Boldfaced values indicate mean recall performance and prediction/effort ratings for the instructions the students found best facilitated recall (upper bold) and that the student would choose again if they were to read another text. Scheffé's procedure was used to test for significance within columns: Subscript *a* indicates significance between READING and SELECTED, subscript *b* indicates significance between READING and GIVEN, and subscript *c* indicates significance between GIVEN and SELECTED ($ps < .05$).

Table 8

Post Hoc Groups' Reliability of Recall Predictions (Predicted With Actual) for Question 3 (Facilitated Recall) and Question 4 (Future Processing Choice) and for the Other Two Instructions That the Students Used

Question/Instructions	*n*	Post hoc groups READING	GIVEN	SELECTED
Facilitated				
READING	34	**.19**	$.52^{**}$	.28
GIVEN	34	.28	**.28**	$.54^{**}$
SELECTED	43	$.42^{**}$	$.31^{*}$	**.28**
Future processing choice				
READING	37	**.10**	$.35^{*}$	.32
GIVEN	30	$.38^{*}$	**.26**	$.51^{**}$
SELECTED	44	$.45^{**}$	$.45^{**}$	**.29**

Note. The boldfaced diagonal shows the correlation coefficients for the personally best instructions. These are surrounded by correlations for the other two instructions. READING = instructions were "read to understand," GIVEN = instructions were "read to understand and use the five given words from the text as support," SELECTED = instructions were "read to understand and select five words of your own from the text as support."

$^{*}p < .05$. $^{**}p < .01$.

answers to Question 4 revealed that the students chose the instructions they felt best facilitated concentration and that were most familiar, thus, the instructions they could best control.

Summary of Phase 3. The post hoc group analyses were based on individual preferences. From this perspective, both recall performances and related ratings were influenced by instructions. The instructions that enhanced recall performance reduced prediction accuracy.

Discussion

The present study was designed to yield overall recall prediction and comprehension calibration accuracies resulting from effortful reading. We assumed that the GIVEN and SELECTED instructions required more active processing, thereby increasing the accuracy of students' ratings. Measures of students' verbal and memory abilities were compared with students' evaluations.

Phase 1 of the experiment confirmed the first hypothesis that the students, regardless of verbal skill, would accurately assess their general verbal and memory abilities (Cull & Zechmeister, 1994; McBride-Chang et al., 1993; Wade & Trathen, 1989). These subjective ratings were correlated with recall performance as well as with the verbal test scores. An interaction indicated that the difference between students of high-verbal-skill and students of either medium- or low-verbal-skill was greater for the comprehension questions than it was for the fluency questions. These empirical distinctions seem appropriate as fluent reading is one, but not the only, requirement for proficient reading comprehension. In some cases (as for the low- and medium-verbal-skill students), reading fluency did not guarantee understanding (Spiro & Myers, 1984). The present study did not contain any specific measure of fluency but showed that high-verbal-skill students read the experimental texts significantly more times, recalled more of the text, and best comprehended the text. Furthermore, Gillström and Rönnberg (1994) found that good readers performed better than poor readers on both verbal and memory tests (i.e., working memory and lexical access).

From a social perspective, the students' ratings of their reading and memory abilities may reflect a necessary awareness of their social position in school (Rueda & Mehan, 1986). Guthrie and Kirsch (1987) argued that the social environment affects students' reading. Poor readers are treated differently than good readers, which makes poor readers aware of how they are viewed by others (Persson, 1994).

Phase 2 of the experiment showed that comprehension calibrations were affected by instructions. In particular, significantly lower comprehension calibrations were associated with the SELECTED instructions. Between 36% and 48% of the students made accurate comprehension calibrations (as defined by a correspondence between actual and calibrated performance).

Neither recall prediction nor recall performance was affected by instructions (Torrance, Thomas, & Robinson, 1993; Wade & Trathen, 1989). For all instructions, 60% or more of the students made accurate recall predictions and postdictions (i.e., within the range of 20%). The data therefore suggest that simple rereading can be as effective as other study techniques for enhancing recall predictions and performance (Haenggi & Perfetti, 1992; Kiewra et al., 1991; Wade & Trathen, 1989).

No significant relations were found between predicted or postdicted recall and recall performance for the high-verbal-skill students, but significant relations were found for the two lower verbal-skill groups (cf. Gillström & Rönnberg, 1994). A closer examination showed that within the verbal-skill groups, accuracy of these relations did not differ much. Between 47% and 85% of the students made accurate recall predictions (i.e., within the range of 20%).

Interpretation of these results must be based on the finding that recall prediction accuracy was the same for high- and low-verbal-skill students. However, the finding that only the low- and medium-verbal-skill students had significant correlations presents a problem for such an interpretation. One solution is to define *good recall predictions* as those that are both reliable and accurate. With this definition, the recall predictions of the high-verbal-skill students were closer to chance, and the lower verbal-skill students were better predictors. In addition, it seems as though the lower verbal-skill students gained from task experience (i.e., recall postdiction accuracy) yet the two upper verbal-skill-level groups did not. As suggested, these results could be explained in terms of degree of automatization of reading, that is, the better the person's reading ability, the less attention he or she requires to complete the reading task (Ackerman, 1990). Thus, skilled readers are no longer aware of the subskills of reading, which makes the perceptual process wholistic in nature (LaBerge & Samuels, 1985). Once a reader's reading reaches this wholistic stage, metacognitive measures may not adequately tap the perceptual process. Metacognitive thinking seems to be most useful in the beginning of and in the development of skillfulness. Davou, Taylor, and Worrall (1991) showed that experts relied heavily on retrieval and pattern recognition, whereas novices relied more on general strategies and thinking ability (cf. Haenggi & Perfetti, 1992; Pokay & Blumenfeld, 1990).

In addition, it is conceivable that the high-verbal-skill students recalled the general meaning of the text rather than the exact word-by-word content. In this vein, Persson (1994) interviewed young readers about their experiences with reading. Good readers had positive self-concepts, considered reading to be fun and important to them, and could draw inferences from the texts and summarize them effectively. The poor readers in Persson's study did not respond in this way. Perhaps in the present study, high-verbal-skill students' recall predictions were more theme than verbatim oriented and perhaps theme-oriented recall prediction is less accurate.

In Phase 3, the students assessed the instructions in terms of ease, comprehensibility, recallability, and future processing choice. Over 50% of the students believed that the READING instructions best facilitated comprehension. However, because high comprehension performance was not associated with the READING instructions, this assessment seems to be incorrect. It is possible that the fewer rereadings and the selection of key words associated with the SELECTED instructions were confusing and produced lower ratings of comprehension. As we argued for the general questions about verbal and memory abilities in Phase 1, it might be that comprehension calibrations are more affected by social awareness and thus that ratings of comprehension are more related to knowledge of verbal-skill level than to actual text comprehension. Glenberg and Epstein (1987) showed that experts tended to calibrate comprehension on the basis of what they should know about a specific topic and not on what they had just read. Moreover, Schneider and Laurion (1993) found that students seemed to base their comprehension calibrations on familiarity with the text topic rather than on comprehension performance. In addition, they found that students had greater problems evaluating how unsure they were compared with how sure they were.

The students assessed ease, recallability, and future processing choice more correctly than they did comprehension. The post hoc analyses that we carried out to validate the students' responses revealed alternative ways to interpret the data. From an individual perspective, it was demonstrated that instructions affected both recall performance and recall predictions and that the students were well aware of these effects (Tables 7 and 8).

The students' elaborations of Question 4 (CHOICE) showed that they had distinct personal preferences regardless of instructions and verbal skill. Some students reported a need to study the whole text and preferred the READING instructions (e.g., "I'm not tied up with a few words—can concentrate on reading"); others preferred SELECTED instructions (e.g., "Although my words might seem strange or irrelevant I have chosen them myself—easier to remember"); and still others preferred GIVEN instructions (e.g.,"I did not have to work that hard when I got some help"). Thus, the students found that certain reading strategies required less effort than others and that they were able to provide reasons for their preferences. Students seem to internalize distinct, built-in ways of text processing, presumably on the basis of cognitive styles (Riding & Sadler-Smith, 1992). Thus, regardless of verbal skills and with sufficient time, all students become skilled readers in the sense that they develop their own strategies and are aware of them. In Phase 2 of the experiment, individual preferences were not considered. Consequently, when using their two, nonpreferred instructions, students had to study in the wrong way, which yielded few instructions main effects for recall and so forth. The personally best instructions were those that resulted in the best recall performance but in poor recall prediction accuracy. Torrance et al. (1993) found that their students were immediately attracted to different types of instructions and suggested that students should be exposed to different instructions rather than given any one type haphazardly. Thus, the personally preferred strategy may become more automatized (Ackerman, 1990) and thus improve performance while reducing control (LaBerge & Samuels, 1985).

We hypothesized that increased activity, involvement, and effort on the part of the reader would result in better performance and accuracy of ratings of performance. However, this was not supported by the data. Also unexpected was the degree to which the patterns of comprehension and recall results diverged, which led us to conclude that one should not generalize the outcome of comprehension to recall or vice versa. Instead, reading for understanding and reading for memorization seem to represent two qualitatively unique mental achievements that are affected differently by the same set of variables. The instructions that we used had an overall effect on students' comprehension calibrations but not on their comprehension performance. Presumably, comprehension tasks should not be time restricted and are affected by task familiarity and social factors. When students were grouped by verbal skill, it seemed as though the instructions had no effect on the recall data. However, when students were grouped according to preferred instructions, recall was better for the preferred instructions. Thus, students were able to identify the reading strategy that worked best for them.

Another important result was that most of the students were aware of their general abilities, but when they were asked for more specific measurements, such as how much of a text would be recalled, the ratings of the lower verbal-skill students were the most accurate. Again, this pattern was not found for comprehension. The GIVEN instructions led to the most accurate number of corresponding calibrated and actual comprehension pairs. These instructions are the most similar to those students use in school (Gillström & Rönnberg, 1994)

An important finding is that most of the rating data revealed verbal-skill-level main effects for recall predictions as well as for comprehension calibrations. The high-verbal-skill students consistently provided the highest (or lowest) ratings, the medium-verbal-skill students consistently provided the middle value ratings, and the low-verbal-skill students consistently provided the lowest (or highest) ratings. With the evidence that recall performance and number of correct answers followed this same pattern, the present results paint a clear picture. Only effort ratings did not differ among verbal-skill levels, which was somewhat unexpected. On the basis of the post hoc analyses, this may be explained by the different preferences for instructions within each verbal-skill group. If one does not consider these individual processing choices, then the effort ratings tend to converge, even within verbal-skill groups (Gillström & Rönnberg, 1994; Maki et al., 1990).

Garner (1987) stated that metacognition refers to stable information about cognition containing information about ourselves, about the tasks, and about the strategies used. These classes of knowledge are supposedly highly interactive. The present data suggest that self-knowledge requires maintenance. Proficiency, in contrast, seems to require little metacognitive awareness. Skilled reading or text processing requires less attention but yields the best performance. Metacognitive ability is most helpful in the development of skillful reading or processing. Once skillfulness is achieved, attention can be deployed elsewhere.

References

Ackerman, P. L. (1990). A correlational analysis of skill specificity: Learning, abilities, and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 883–901.

Begg, I. M., Martin, L. A., & Needham, D. R. (1992). Memory monitoring: How useful is self-knowledge about memory? *European Journal of Cognitive Psychology, 4,* 195–218.

Cull, W. L., & Zechmeister, E. B. (1994). The learning ability paradox in adult metamemory research: Where are the metamemory differences between good and poor learners? *Memory & Cognition, 22,* 249–257.

Davou, B., Taylor, F., & Worrall, N. (1991). The interplay of knowledge and abilities in the processing of text. *British Journal of Psychology, 61,* 310–318.

Garner, R. (1987). *Metacognition and reading comprehension.* Norwood, NJ: Ablex.

Gillström, Å., & Rönnberg, J. (1994). Prediction accuracy of text recall: Ease, effort, and familiarity. *Scandinavian Journal of Psychology, 35,* 367–386.

Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition, 15,* 84–93.

Guthrie, J. T., & Kirsch, I. S. (1987). Distinctions between reading comprehension and locating information in text. *Journal of Educational Psychology, 79,* 220–227.

Haenggi, D., & Perfetti, C. A. (1992). Individual differences in reprocessing of text. *Journal of Educational Psychology, 84,* 182–192.

Kiewra, K. A., Mayer, R. E., Christensen, M., Sung-Il, K., & Risch, N. (1991). Effects of repetition on recall and note-taking: Strategies for learning from lectures. *Journal of Educational Psychology, 83,* 120–123.

Kirk, R. E. (1968). *Experimental design: Procedures for the behavioral sciences.* Belmont, CA: Brooks/Cole.

LaBerge, D., & Samuels, S. J. (1985). Toward a theory of automatic information processing in reading. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading (3rd ed.).* Newark, DE: International Reading Association.

Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 663–679.

Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 609–616.

Maki, R. H., & Serra, M. (1992). Role of practice tests in the accuracy of test predictions on text material. *Journal of Educational Psychology, 84,* 200–210.

Maki, R. H., & Swett, S. (1987). Metamemory for narrative text. *Memory & Cognition, 15,* 72–83.

Marascuilo, L. A., & Levin, J. R. (1970). Appropriate post hoc comparisons for interaction and nested hypotheses in analysis of variance designs. *American Educational Research Journal, 7,* 397–421.

Mazzoni, G., & Cornoldi, C. (1993). Strategies in study time allocation: Why is study time sometimes not effective? *Journal of Experimental Psychology: General, 122,* 47–60.

McBride-Chang, C., Manis, F. R., Seidenberg, M. S., Custodio, R. G., & Doi, L. M. (1993). Print exposure as a predictor of word reading and reading comprehension in disabled and nondisabled readers. *Journal of Educational Psychology, 85,* 230–238.

McDaniel, M. A. (1984). The role of elaborative and schema processes in story memory. *Memory & Cognition, 12,* 46–51.

Necka, E., Machera, M., & Miklas, E. (1992). Incidental learning, intelligence, and verbal ability. *Learning and Instruction, 2,* 141–153.

Noice, H. (1993). Effects of rote versus gist strategy on the verbatim retention of theatrical scripts. *Applied Cognitive Psychology, 7,* 75–84.

Otero, J., Campanario, J. M., & Hopkins, K. D. (1992). The relationship between academic achievement and metacognitive comprehension monitoring ability of Spanish secondary school students. *Educational and Psychological Measurement, 52,* 419–430.

Persson, U.-B. (1994). *Reading for understanding: An empirical contribution to the metacognition of reading comprehension* (Linköping Studies in Education and Psychology No. 41). Linköping, Sweden: Linköping University, Department of Education and Psychology.

Pokay, P., & Blumenfeld, P. C. (1990). Predicting achievement early and late in the semester: The role of motivation and use of learning strategies. *Journal of Educational Psychology, 82,* 41–50.

Pressley, M., Snyder, B. L., Levin, J. R., Murray, H. G., & Ghatala, E. S. (1987). Perceived readiness for examination performance (PREP) produced by initial reading of text and text containing adjunct questions. *Reading Research Quarterly, 22,* 219–236.

Psykologiförlaget. (1976). *Manual till Diagnostiska Läs-och Skrivprov för Högstadiet.* [Manual to Diagnostic Reading-and-Writing Test for Students Grade 7 to 9: Senior Level of Compulsory School]. Stockholm: Psykologiförlaget.

Riding, R., & Sadler-Smith, E. (1992). Type of instructional material, cognitive style and learning performance. *Educational Studies, 18,* 323–339.

Rueda, R., & Mehan, H. (1986). Metacognition and passing:

Strategic interactions in the lives of students with learning disabilities. *Anthropology & Education Quarterly, 17,* 145–165.

Schneider, S. L., & Laurion, S. K. (1993). Do we know what we've learned from listening to the news? *Memory & Cognition, 21,* 198–209.

Spiro, R. J., & Myers, A. (1984). Individual differences and underlying cognitive processes in reading. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 471–501). New York: Longman.

Torrance, M., Thomas, G. V., & Robinson, E. J. (1993). Training in thesis writing: an evaluation of three conceptual orientations. *British Journal of Educational Psychology, 63,* 170–184.

Wade, S. E., & Trathen, W. (1989). Effect on self-selected study methods on learning. *Journal of Educational Psychology, 81,* 40–47.

Westrin, P. A. (1965). *WIT III manual.* Stockholm: Psykologiförlaget.

Appendix

Three Experimental Texts, Supportive Words, and Accompanying Comprehension Questions for Each Text

The text examples are from *Manual till Diagnostika Läs-och Skrivprov för Högstadiet* [Manual to Diagnostic Reading-and-Writing Test for Students Grade 7 to 9: Senior Level of Compulsory School] (pp. 2–17), by Psykologiförlaget, 1976, Stockholm: Psykologiförlaget. Copyright 1976 by Psykologiförlaget. Adapted with permission. Correct answers are indicated by asterisks.

Text 1: Banana

You probably find the banana a natural product in the supermarket and in the kiosk. In the beginning of the 20th century, however, it was almost an unknown fruit in Europe. The banana was one of the first plants cultivated by the people in East Asia, the original homeland of the banana. When Alexander the Great invaded India in the year 327 B.C., his armies found lots of banana plants by the riverside of the Indus. Presumably, it was then discovered that the dried roots easily could be transported and planted in other hot and humid areas of the world, where good soil was found. The plant immediately put forth new sprouts, spread its large leaves, flourished and bore fruit—all with amazing speed. Eventually the banana plant dispersed to Africa as well as Australia and the Pacific. A Spanish monk introduced the banana in America shortly after Columbus had discovered the Antilles.

Given key words: *banana, East Asia, Alexander, dispersion,* and *monk.*

Comprehension Question 1. How is the content of the text best summarized?

a. India—the homeland of the banana.
*b. The banana and its dispersion.
c. The world's fastest growing tree.
d. The banana's way to America.

Comprehension Question 2. The banana plant is today found in many different places on earth mainly because

*a. the roots can easily be transported and planted elsewhere.
b. it grows quickly and bears fruit.
c. it can grow anywhere in hot areas.
d. Alexander's armies dispersed it.
e. it can grow anywhere where the soil is good.

Text 2: Arabia

The sun suddenly rose on the slope where the seven Arab tents were located deep inside the great deserts of Arabia. The tents stood in a half circle, below the sandhills, and in front of every tent rose a blue smoke pillar from a fire of camel dung, where Arabian women with veils thrown back from their faces prepared rice and bread to break the long fast of the night. Around the largest tent sat a group of men on their heels, drawing circles in the sand with their shepherd's sticks or drinking bitter Arabic coffee out of small earless cups. This tent belonged to the leader of the group, and the men in the camp were discussing what they should do. All were barefoot and wore black or brown woollen cloaks over a loose dress of white cloth. Their heads were covered by red and white checked cloths, trimmed with tufts, fixed with black woollen ribbons.

Given key words: *Arab tents, smoke pillar, shepherd's sticks, leader,* and *discussion.*

Comprehension Question 1. Which statement is most correct?

a. The Arab women prepared coffee, bread and rice.
b. In front of the largest tent women prepared rice and bread.
c. Women with veils in front of their faces prepared rice.
*d. The men sat around the fire and drank coffee.
e. The men sat by the opening of the tent and drew in the sand.

Comprehension Question 2. Which headline is the best for this paragraph?

*a. Morning in an Arab camp.
b. Sunrise in the desert.
c. Around the fires in the desert.
d. One day among the Arab people.
e. Desert life in Arabia.

Text 3: Indian

The first time a Westerner hears an Indian sing and play music he probably feels very confused, because the music sounds totally different from all he has ever heard. The sound seems harsh, the notes are gliding up and down in a way which never would be accepted in Europe or America. But the Easterner might also be as confused when he hears Western music. He thinks our 12-tone scale sounds false. One of the most striking differences between Eastern and Western music is found in the way it is performed. In a Western orchestra the musicians sit in front of the listeners on a bandstand. In India the musicians sit with their legs crossed on the small part of the floor not occupied by the listeners, who also sit

(*Appendix continues on next page*)

Appendix (Continued)

with their legs crossed on the floor in a half circle. There is no applause, which is regarded "barbaric."

Given key words: *Indian*, *false*, *perform*, *sit*, and *applause*.

Comprehension Question 1. The best headline for this paragraph is
a. A Westerner listens to Eastern music.
b. Sounds and notes in India and Europe.
*c. Differences between Eastern and Western music.
d. How one sings and plays music in India and America.
e. Music from different parts of the world.

Comprehension Question 2. What do Westerners find strange when Eastern music is performed?
a. The notes, which seem false.
*b. The notes, which are gliding up and down.
c. The music, which sounds barbaric.
d. The scale, which has twelve tones.
e. The sound, which seems Eastern.

Received August 16, 1993
Revision received June 13, 1995
Accepted June 16, 1995 ■

Studie III

Comprehension Calibration and Memory Prediction Accuracy of Texts: The effect of reading to remember and reading to comprehend instructions on performance and accuracy of ratings

Åsa Eriksson
Department of Education and Psychology
Linköping University
Sweden

Address for correspondence: Åsa Eriksson, Department of Education and Psychology, Linköping University, S-581 83 Linköping, Sweden. E-mail: asaer@ipp.liu.se.

RUNNING HEAD: Read to remember and understand.

Abstract

The present study investigated students' metacomprehension and metamemory of texts. High-school students read two texts, one in order to remember as much as possible and another to comprehend it as well as possible. The students calibrated their comprehension and predicted memory of these texts. They were tested immediately and after a delay of one week, half of the students were given four minutes to read the texts whereas the other half were given free reading time. It was assumed that concentrating on one processing task at a time should improve performance and accuracy of ratings. No calibration accuracy was received whereas the students demonstrated immediate postcalibration accuracy. The students could accurately predict their immediate but not delayed text recall. Immediate and delayed postdiction accuracy was found indicated a well-kept conception of the text to related to even after a delay. The lower verbal skill students made the most reliable predictions, whereas the high verbal skill students excelled in performance. The students were able to evaluate which instruction best facilitated their text recall. Reading time per se had no effect on performance or accuracy of ratings.

This study is concerned with memory and comprehension monitoring in terms of performance predictions of comprehension and memory of texts. If students' ratings of their understanding of texts correspond with actual answering performance on comprehension questions they show calibration accuracy. Similarly, if students' ratings of what they remember of texts correspond with actual recall performance they show prediction accuracy (Eriksson & Rönnberg, 2000; Gillström & Rönnberg, 1995; Hacker, 1998; Lin & Zabrucky, 1998; Maki & Berry, 1984; Pressley & Schneider, 1997).

Lin and Zabrucky (1998) concluded that the calibration accuracy studies have demonstrated low and sometimes, insignificant correlations. They suggested that this metacomprehension measure is sensitive towards various methodological changes, which our previous studies also confirm. Gillström and Rönnberg (1995) found calibration accuracy whereas Eriksson and Rönnberg (2000) did not. The fact that there were twice as many students in the former study could have led to a better range of calibrations than in the latter. Also, Hallam and Francis (1998) found that experienced readers' understanding varied both between and within different texts presumably due to interest and prior knowledge. The present study investigated how text interest affected comprehension and text recall.

Pressley and Schneider (1997) suggest that the logic behind prediction accuracy is that if people have monitored previous experiences they should also be able to predict future performance accurately and thereby show memory monitoring. Pressley and Schneider (1997) studied over a hundred metamemory studies and found an average correlation coefficient between predicted and actual memory performance of r. 41. This suggested level of statistical association has also been found in our previous studies, r´ s between .30 and .55 (Gillström & Rönnberg, 1994, 1995; Eriksson & Rönnberg, 2000). Thus, the present students were expected to accurately predict their immediate recall performance.

Usually metacomprehension and metamemory are studied alone but this study combined both aspects to present a broader view on learning. As a learner, the students have to be able to both evaluate their comprehension and memory of text (Eriksson &

Rönnberg, 2000; Gillström & Rönnberg, 1995; Lovett & Pillow, 1995, 1996). Research has focused on subject- (e.g. skill), task- (e.g. instructions, immediate/delay) and text-related factors (e.g. genre), and how these affect self-awareness (Lin & Zabrucky, 1998; Pressley & Schneider, 1997). Our metamemory data have shown that high verbal skill students better recall texts but make less accurate predictions of performance than lower verbal skill students (cf. Maki & Swett, 1987). Especially their retrospective ratings (i.e., postdictions) have been more accurate. One reason could be that the reading in former groups is more attention-free, automatic and writing down what they remember constitutes a rather effortless task resulting in less metacognitive awareness (Eriksson & Rönnberg, 2000; Gillström & Rönnberg, 1994, 1995; Logan, 1988). From a metacomprehension point of view, no verbal skill differences have been found (cf. Pressley, Snyder, Levin, Murray & Ghatala, 1987).

Pressley and Schneider (1997) suggested that the key to become efficient information processors is to use instructions that show learners what to do. Gillström and Rönnberg (1994) showed that students found learning school-book texts more familiar and effort requiring than reading fairy-tales trying to teach someone the contents. The former resulted in immediate as well as delayed prediction accuracy whereas the latter did not. In this vein, McDaniel (1984) found reading texts with deleted letters required more effort than intact texts resulting in better text recall. Maki, Foley, Kajer, Thompson and Willert (1990) showed that these deleted letter texts resulted in better accuracy of performance ratings as well.

In a later study, Gillström and Rönnberg (1995) found that when students studied texts in a preferred way they could identify which instruction resulted in best level of text recall. This instruction also resulted in lower prediction accuracy implying a similar pattern of data as previously described for metamemory and verbal skill. This study preference effect was not found for metacomprehension which is one of the reasons for this study. One of the reasons could be rivalry between the goals of text processing. The students were instructed to read to comprehend and use no, given, or selected key words as additional help. The students were then tested on both their comprehension and memory of texts. The present study let the students read two texts, one with the instruction to remember as much as possible, the other to comprehend it as well as they could. It was assumed that focus on one mental process at a time should have a positive effect on ratings of performance for both (Lovett & Pillow, 1995, 1996). If the students were able to follow instructions, it should show via best possible comprehension with the 'understand' instruction and best possible recall with the 'remember' instruction. Pressley and Schneider (1997) argued that within-subject designs is needed to find out if students can adjust to instruction requirements. Via this design, Gillström and Rönnberg (1995) showed the importance of personal preferences.

Successful learning requires planing for upcoming tests and homework, which makes it necessary to study both immediate and long-term metacognition (Hacker, 1998; Lin & Zabrucky, 1998). Gillström and Rönnberg (1994) used a week's delay between reading and prediction accuracy and found a reduction in both performance and metamemory with one exception. Those instructed to learn a school-book text demonstrated delayed prediction accuracy presumably due to high effort demands and levels of familiarity (cf. Maki & Swett, 1987). Eriksson and Rönnberg (2000) introduced a delay of one month and found a clear reduction in both performance and calibration and prediction accuracies. However, postdiction accuracy of texts was significant even after a month delay. The students knew how well they managed to recall one month after having read the text, indicating that these students had a well kept conception of the text to relate to even after such a long delay (Eriksson & Rönnberg, 2000). In this study, a week's delay was used and the students made both prospective and retrospective ratings of performance. The delay should have a negative effect on performance as well as performance predictions with the exception of retrospective ratings.

As a control purpose, half of the present students planned their own reading time, the other half was allowed four minutes reading. Mazzoni and Cornoldi (1993) concluded that extended reading time per se did not increase text recall performance. Gillström and Rönnberg (1994) found that free disposal of reading time in combination with easy text materials made the students spend too little time reading the text to be able to remember and make accurate ratings. Cull and Zechmeister (1994) found that their students did not study test items long enough even when study time was unlimited. Also, earlier studies have let students rate their general reading and memory abilities. These ratings have correlated positively with actual text recall and comprehension and also with verbal tests result which has been taken as an evidence of internal validity. The present students also made these ratings.

To summarize: The present study made an attempt to improve calibration accuracy and maintain prediction accuracy. The students were instructed to concentrate on one mental process at a time - reading to comprehend and then reading to remember. A week's delay was expected to reduce performances as well as accuracy of ratings but free or four-minute reading times were not (Cull & Zechmeister, 1994; Gillström & Rönnberg, 1994, 1995; Eriksson & Rönnberg, 2000; Mazzoni & Cornoldi, 1993). The students' interest ratings in texts should correlate positively with actual memory and level of comprehension of texts (Hallam & Francis, 1998). The answer to open-ended questions on students' views on reading for comprehension and memory are reported in the Discussion to complement the quantitative data with qualitative reports.

Method

Participants

A total of 88 high-school students participated in the immediate testing. There were four classes selected from the same school. Two of these classes included students from vocational training programs and two students from theoretical programs of social science. Their mean age was 17.56 years (SD .63). Of these, 68 participated in the delayed testing after one week.

Verbal Material

Two short expository texts were used as reading materials (see Appendix 1). The texts were taken from a standardized diagnostic test battery consisting of five separate tests, which are used in Sweden for the assessment of reading and writing achievements of students in Grades 7-9 (Psykologiförlaget, 1976). One of the tests consists of 14 short expository texts accompanied by 30 multiple-choice questions assessing reading comprehension. The selection of the two experimental texts was made in a pilot study (Gillström & Rönnberg, 1995). A group of students read the 14 texts and answered the questions. Those two texts that received approximately 50 percentage correct answers were chosen as experimental texts (Gillström & Rönnberg, 1995).

Reading Instructions

One text was read in order to COMPREHEND it as good as possible, the other to REMEMBER as much as possible. Half of the students had no reading time-limit (i.e., FREE), half read the text for exactly four minutes (i.e., FOUR; Gillström & Rönnberg, 1995; Eriksson & Rönnberg, 2000):

1.1 Read to understand (COMPREHEND): You will now read the text through until you feel that you have understood what it is all about. After 4 minutes you are going to answer two questions concerning your comprehension of the text, and you will also try to recall as much of the text as possible (FOUR).

1.2 Read to understand (COMPREHEND): You will now read the text through until you feel that you have understood what it is all about. Use the time you need. Write down what time you started and when you stopped reading the text. After that you are going to answer two questions concerning your comprehension of the text, and you will also try to recall as much of the text as possible (FREE).

2.1 Read to remember (REMEMBER): You will now read the text through with the purpose to remember as much as possible. After 4 minutes you are going to answer two questions concerning your comprehension of the text, and you will also try to recall as much of the text as possible (FOUR).

2.2 Read to remember (REMEMBER): You will now read the text through with the purpose to remember as much as possible. Use the time you need. Write down what time you started and when you stopped reading the text. After that you are going to answer two questions concerning your comprehension of the text, and you will also try to recall as much of the text as possible (FREE).

Design

A repeated measures design was used with instruction (REMEMBER and COMPREHEND) and time of testing (IMMEDIATE and DELAYED) as within subjects variables, and verbal skill (HIGH; MEDIUM and LOW) and reading time (FREE or FOUR) as between subjects variables. Order of instructions and texts was balanced across students (cf. Gillström & Rönnberg, 1994, 1995, Eriksson & Rönnberg, 2000).

General Design and Procedure

The students were tested twice. Immediately and after a week' s delay. Both times the experiment took place at their school in one of the classrooms during Swedish lessons. All test materials were presented in the classroom setting. Approximately 15 to 20 students participated at the same time. Beforehand, the teachers were instructed to divide the students into two equal groups of verbal skill (Gillström & Rönnberg, 1994; Necka, Machera & Miklas, 1992). Table 1 presents the overall, serial chain of events with the exact wording of instructions and questions. Below follows a general description of the study.

Immediate testing. The students were given a booklet, which contained all experimental material. First, the students rated their general reading and memory abilities and completed the synonym/antonym and analogy verbal tests. After that, the FREE-condition students went to a near-by classroom where they worked the rest of the booklet through at their own pace. The FOUR-condition students also worked the booklet through but with reading time restrictions.

Insert Table 1 about here

The students rated[1] their general fluency of reading, their reading comprehension and their ability to memorize texts. In the Results section below the verbal test scores are reported together with ratings of general ability (Table 1).

---

[1] Throughout the experiment the same type of ratings scale was used: The students could mark their ratings anywhere on a 10-cm scale. The students were told to interpret the scale in terms of percentage.

The verbal tests, antonym/synonym and analogies, belonged to a standardized test battery used in Sweden to measure study success (Westrin, 1965). For ninth graders the average score on the synonym-antonym test is 14.70 out of 29. As a practice example, the students were presented with the following five words: false, rare, erroneous, genuine, and whole. The students were to mark the two words that are opposite in meaning. They were given five minutes to complete the test. The analogy test measures verbal inductive ability. As a practice example, the students were presented with the word pair driver-car and were instructed to find an analogous pair out of the following five words: trot, riding, horse, ride and rider. For ninth graders the average score on the analogy test is 15.5 out of 27. The students were given five and a half minute to complete the test. The average score on both these tests, combined, was 15.1 and the correlation between the test scores was *r* = .71, for ninth-graders.

Reading task. FOUR or FREE reading of text was followed by ratings of performance, effort and interest (Table 1). Actual comprehension was measured via two multiple-choice questions (see Appendix 1). Actual test of text recall consisted of the students trying to recall as much as they remembered. They were instructed to write down "everything" they remembered, even if they were uncertain as to the exact words or order of appearance. The students made four post-ratings of performance (Table 1). Finally, after having read the two texts the students evaluated both instructions in terms of preference for recall and comprehension. Before ending the immediate testing, the students predicted how much they would recall and how well they would comprehend the text in a week's time (Table 1).

At the delayed testing the students were tested on their memory and comprehension of texts a second time. They answered open-ended questions regarding their way of working with memory prediction and comprehension calibration accuracy, and also what they think constitutes good readers, how they remember and comprehend texts, and if they think that these tasks are similar or different tasks (Gillström & Rönnberg, 1995). After having completed the experiment the students were debriefed about the purpose of the study.

Results

Our initial analyses showed that FREE or FOUR minute reading groups were not different in terms of verbal test scores, performance, ratings, or accuracy of ratings, with two exceptions. A significant instruction by number-of-readings interaction indicated that the FOUR-readers read the texts more times than the FREE-readers (3.74, REMEMBER and 3.63, COMPREHEND), and that FREE-readers read the texts more times with REMEMBER (1.98) than COMPREHEND (1.30), 3.63, $\underline{F}(1, 75) = 5.67$, $\underline{p} < .05$, $\underline{MSE} = .60$. A main effect of reading time indicated that the FREE-readers spent half the time reading the texts (≤ 2.00 minutes) compared to the FOUR minute readers, $\underline{F}(1, 69) = 292.54$, $\underline{p} < .05$, $\underline{MSE} = .41$. Thus, those who planned their own reading time (FREE) spent less time reading the texts and they read the texts fewer times compared with the FOUR-readers. Since these differences had no affect on the rest of the data, reading time groups were collapsed into one.

Analyses of data are reported overall for each instruction, for time of testing (see Method) and by verbal skill. The 20% top and bottom scorers on the verbal tests were regarded HIGH and LOW, respectively, compared with the larger MEDIUM group (Eriksson & Rönnberg, 1999; Gillström & Rönnberg, 1995). As a validation of verbal skill levels, verbal test results should correlate positively with text recall, answering performance, and with ratings of general verbal and memory abilities (Eriksson & Rönnberg, 2000; Gillström & Rönnberg, 1994, 1995).

Part 1 of the result section reports the general ratings of verbal ability together with the verbal test scores. Part 2a) report data related to calibration accuracy 2b) data related to prediction accuracy.

Part 1: General ratings of verbal ability and verbal test scores

The overall mean score on the word knowledge test (i.e., the synonym/antonym test) was 15.24 out of 29 ($\underline{SD} = 4.77$) and on the analogy test 15.23 out of 27 ($\underline{SD} = 3.06$). The correlation coefficient between the tests was $\underline{r} = .51$, $\underline{p} < 01$. These means and the correlation coefficient correspond well with the standardized results in the WIT-

III manual from which these tests were taken (Eriksson & Rönnberg, 2000; Gillström & Rönnberg, 1995; Westrin, 1965).

Insert Table 2 about here

A two-factor ANOVA on the general questions with verbal skill as a between-subjects factor, and questions as within-subject factor, demonstrated a significant general questions effect, $F(2,172) = 11.63$ $p < .01$, $MSE = 192.33$ No verbal skill differences, $F(2, 84) = 2.08$, $p > .05$, $MSE = 397.04$, or interaction were found, $F(4,168) = 1.00$, $p > .05$, $MSE = 192.33$. Regardless of verbal skill, data suggest that students rated their memory for texts as lower compared to their fluency and comprehension, t' s 4.44 and 4.73, p' s $< .01$ (Table 2) (cf. Eriksson & Rönnberg, 2000). Although not significant, there were tendencies indicating that the high verbal skill group made higher ratings than the lower verbal skill groups (Gillström & Rönnberg, 1995; Eriksson & Rönnberg, 2000).

Insert Table 3 about here

Table 3 shows that the verbal tests could be used to divide the students into verbal skill groups in that both objective and subjective aspects of comprehension and memory of texts correlate with each other (Gillström & Rönnberg, 1994, 1995: Eriksson & Rönnberg, 2000; Stark, Renkl, Gruber & Mandl, 1998).

Data related to calibration accuracy of comprehension

Comprehension calibrations. The students made both immediate and delayed calibrations of level of comprehension for each instruction. A three-factor ANOVA on these calibrations revealed significant main effects of verbal skill , $F(2, 80) = 9.64$, $p < .01$, $MSE = 570.01$, and time of testing, $F(3, 243) = 106.43$, $p < .01$, $MSE = 285.35$, but not of instruction, $F < 1$. A significant verbal skill by instruction by time of testing interaction indicated that LOW achievers expected poorer comprehension with

COMPREHEND than REMEMBER whereas both HIGH and MEDIUM performers expected the opposite, $F(6, 243) = 2.77$ $p < .05$, *MSE* = 188.73 (Table 4).

Postcalibrations. The students made two postcalibrations concerning the number of questions they believed they had answered correctly. These were made directly after having answered the questions at immediate and then at delayed testing. A three-factor ANOVA on these comprehension postcalibrations revealed a significant verbal skill, $F(2, 60) = 3.06$, $p = .05$, *MSE* = .17, and time of testing main effects, $F(2, 52) = 9.71$, $p < .05$, *MSE* = .04. No instruction effect or interactions were detected, F' s between .00 and 2.12 (Table 4). Across instruction and time of testing, the HIGH made higher comprehension postcalibrations (.68) than the LOW (.47), MEDIUM (.63), $p < .05$.

Actual comprehension. The students answered two multiple-choice questions following each text twice, immediately and after the delay (Gillström & Rönnberg, 1994, 1995, Eriksson & Rönnberg, 2000). A three-factor ANOVA on answering performance, with verbal skill as a between-subject factor and instructions and time of testing as within-subjects factors revealed a main effect of verbal skill, $F(2, 63) = 3.87$, $p < .05$, *MSE* = .21. Across instructions and time of testing, mean level of comprehension was (.66) for HIGH, (.63) for MEDIUM and (.44) for LOW, t' s 2.29 and 2.66 respectively, $p < .05$, between the higher and the low verbal skill students (Table 4).

Insert Table 4 about here

Calibration accuracy of comprehension. As shown in Table 5, a clear pattern of insignificant correlations was obtained between the comprehension calibrations and level of comprehension (Eriksson & Rönnberg, 2000). The comprehension postcalibrations increased the reliability for immediate but not delayed data and the latter result could be due to the students' ratings being significantly lower but level of comprehension remained the same after a week (Table 5).

The correlation coefficient is an indication of reliability and to assess actual accuracy the difference between ratings of comprehension and actual answering performance was calculated. The smaller the difference the better calibration accuracy (Eriksson & Rönnberg, 2000; Gillström & Rönnberg, 1995). These analyses did not reveal anything out of the ordinary. Up to 50 percent of the students made inaccurate ratings at the immediate testing whereas up to 65% did so at the delayed testing.

Students' assessments of best instruction for comprehension. The present students evaluated which instruction best facilitated comprehension (Gillström & Rönnberg, 1995; Eriksson & Rönnberg, 2000). A three-factor ANOVA with best instruction for level of comprehension as between-subjects factor, instruction and time of testing as within-subjects factors did not reveal any significant interactions for best instruction for comprehension, *F*'s between .38 and 2.16, $p > .05$ (Table 8) (Gillström & Rönnberg, 1995; Eriksson & Rönnberg, 2000). Thus, the students could not identify which instruction made them answer the questions most correctly.

Insert Table 5 about here

Data related to prediction accuracy of text recall

Memory predictions of text recall. The students both made immediate and delayed predictions of text recall. Only time of testing revealed a significant main effect for memory predictions indicating that the students expected to recall less in a week's time, $F(1, 81) = 102.99$, $p < .01$, $MSE = 195.70$. There was a tendency towards significance for verbal skill, $F(2, 81) = 2.58$, $p = .08$, $MSE = 514.74$. No main effect of instruction or interactions were detected, *F*'s between .09 and 1.69 (Table 6).

Memory postdictions of text recall. The students both made immediate and delayed postdictions of text recall. A three-factor ANOVA revealed that postdictions did not differ for verbal skill, $F(2, 59) = 2.28$, $p > .05$, $MSE = 881.40$, or by instructions, $F(2, 59) = 1.53$, $p > .05$, $MSE = 295.65$, but by time of testing indicating that the students made higher immediate than delayed postdictions, F(1, 59) = 25.77, p < .01, MSE = 191.80. No interactions were detected, F's between .22 and .83 (Table 6).

Recall performance. Each of the two texts was divided into 33 propositions. Text recall was scored in a deviation from verbatim fashion; correct recall could include any of the following nonessential deviations from true verbatim; adding conjunctions, such as and, change form of verb, and substituting singular form for plural (for further details see Noice, 1993).

A three-factor ANOVA on recall revealed significant effects of instruction showing that the students recalled more with REMEMBER than COMPREHEND, F(1, 64) = 4.00, p = .05, MSE = .04, and time of testing indicating a poorer recall after a week's time, F(1, 64) = 83.25, p < .04, MSE = .02. A verbal skill main effect indicate that HIGH (.39) and MEDIUM (.33) recalled more than LOW (.23), t' s 2.58 and 3.40, F(2, 64) = 4.82, p < .05, MSE = .06. No interactions were detected, F's between .22 and 2.12 (Table 6).

Insert Table 6 about here

Memory prediction accuracy. The students made reliable memory predictions of immediate but not delayed recall performance, hence replicating previous findings (Eriksson & Rönnberg, 2000; Gillström & Rönnberg, 1994; Gillström & Rönnberg, 1995). As expected the LOW and the MEDIUM students made the most reliable ratings of performance (Table 7).

Memory postdiction accuracy. A clear pattern of overall and also lower verbal skill significant correlations was obtained for the immediate and delayed memory postdictions (Eriksson & Rönnberg, 2000; Gillström & Rönnberg, 1995; Maki, 1998). The HIGH demonstrated that they could reliably memory postdict their immediate REMEMBER recall (Table 7).

Insert Table 7 about here

To study accuracy of ratings the difference between memory predicted/postdicted and actual recall should be as low as possible, maximum ±20% (Gillström & Rönnberg, 1995; Eriksson & Rönnberg, 2000). Between 65 and 80% of the students' immediate or delayed ratings fell within the acceptable range. Most of the unacceptable ratings were overestimations (up to 29%).

Students' assessments of best instruction for recall. A three-factor ANOVA evaluation of facilitation of text recall, revealed a two-factor interaction between preferred instruction and recall, $F(2, 58) = 11.09$, $p < .05$ (Gillström & Rönnberg, 1995; Eriksson & Rönnberg, 2000). Thus, the students knew which instruction improved their recall - those who claimed better recall with immediate REMEMBER recalled 48% with that instruction compared to 31% for COMPREHEND, and so forth (Table 8) (Eriksson & Rönnberg, 2000; Gillström & Rönnberg, 1995).

Ratings of interest. A two-factor ANOVA on text interest did not differ due to verbal skill or instruction, *F*'s between .48 and 1.92. Overall mean ratings with COMPREHEND was 40.74 and with REMEMBER 35.96. Ratings of interest correlated with immediate recall for both COMPREHEND and REMEMBER, *r*' s .27 and .22, $p < .05$ respectively, indicating that the more interesting texts the better recall. Level of comprehension did not correlate with interest for either instruction.

Discussion

This study investigated comprehension and memory monitoring in terms of performance predictions of texts. Previous studies indicated that high-school students better can evaluate their memory than comprehension of texts (Gillström & Rönnberg, 1995; Eriksson & Rönnberg, 2000). Therefore, the present study set out to improve students' ratings of comprehension via the use of reading instructions and free report of comprehension. Unfortunately, none of these attempts turned out successful. Reading to remember improved text recall but read to understand did not improve level of comprehension in a similar way. In addition, the students showed prediction but not calibration accuracy. The introduction of free-report of comprehension did not work in an intended way (Koriat & Goldsmith, 1997). It seemed as if some of the students mistook this task with text recall, that is, confused comprehension of text with memory of text.

A control purpose was to study the effect of reading time. Those students who were given FREE reading read the texts fewer times and spent half the time reading the texts compared to the FOUR minute readers but this had no affect on performance or performance predictions (Gillström & Rönnberg, 1994, 1995; Eriksson & Rönnberg, 2000, Mazzoni & Cornoldi, 1993). Since measures of verbal ability indicated that the FREE and FOUR groups were equal in terms of word knowledge and verbal inductive ability, the result of the time groups were collapsed.

Even with these new instructions many previously attained data patterns were replicated. The students predicted immediate but not delayed memory performance accurately. Both immediate and delayed postdiction accuracy indicated that the students had a well kept conception of the text to which they related to. Postcalibrations of the number of questions answered correctly were also made rather accurately whereas overall comprehension calibrations do not match the level of comprehension as accurately ((Eriksson & Rönnberg, 2000; Gillström & Rönnberg, 1994; 1995; Maki, 1998). The lower verbal skill groups made the more accurate memory prediction and postdiction ratings and the students demonstrated study preferences for memory but not for comprehension of text. That is, the students knew which of the two instructions make them recall the most, but this is not applicable to comprehension (Gillström & Rönnberg, 1995; Eriksson & Rönnberg, 2000).

Performance predictions of comprehension remains a complicated matter (Lin & Zabrucky, 1998). Benjamin, Bjork and Schwartz (1998) found that students sometimes fail on tests even when they feel ready, which could be due to their criteria of learning not matching with the task demands at hand (cf. Wenestam, 1993). Conway, Gardiner, Perfect, Anderson and Cohen (1997) described the learning process during a university course. Only after reading and discussion students learn and are able to use the information. In this vein, Eriksson and Rönnberg (2000) found a clear reduction in recall performance in a month's time, whereas level of comprehension remained quite intact. Some of their students even answered the questions as good as or better after the delay, but they themselves expected a reduction (Gillström & Rönnberg, 1994; Conway, et al., 1997).

Some of the present data point in the direction that reading to comprehend is regarded as an easier, less effort requiring task than reading to remember. When some of the students described their way of going through the experimental tasks, they indicated that they spent more time and effort remembering texts compared to understanding them: "To read with the purpose to understand is much easier.", "The one I read to remember I read much more carefully.". Furthermore, one of the open-ended questions addressed the question of how accurately the students thought their ratings were. More than 60% of the students indicated that they were satisfied with their comprehension calibrations (e.g. "There is pretty good agreement"., "Good!") (cf. Wenestam, 1993; Benjamin et. al., 1998), whereas less than 50% were satisfied with their memory predictions (e.g. "A bit too high"., "Not so good!"). These answers are contradictory to our findings and could be based on the fact that the students are familiar with the topics of the texts – Bananas, Arabic desert (Hallam & Francis, 1998). Schneider and Laurion (1993) found that comprehension calibrations are based on familiarity with the topic rather than on actual content, hence suggesting an illusion-of-knowing effect. Thus, experts rely on prior knowledge when they read texts within their own field as their ratings of performance were no more accurate than those of novices (Glenberg & Epstein, 1987).

In closing, students accurately predict their memory of texts but they have problems calibrating their comprehension thereof. Our data suggest that effort requirements is one of those factors that could explain why this is the case. When a reader has to pay close attention he/she can more accurately evaluate their performance. Future research will show if this hypothesis remains true.

References

Anderson, R. C., & Freebody, P. (1985). Vocabulary knowledge. In H. Singer & R. B. Ruddell (Eds.), Theoretical models and processes of reading, third edition. Newark: IRA.

Benjamin, A. S., Bjork, R. A., & Bennett, L. S. (1998). The mismeasure of memory: when retrieval fluency is misleading as a metamnemonic index. Journal of Experimental Psychology: General, 127, 55-68.

Conway, M. A., Gardiner, J. M., Perfect, T.J., Anderson, S.J., & Cohen, G. M. (1997). Changes in memory awareness during learning: The acquisition of knowledge by psychology undergraduates. Journal of Experimental Psychology: General, 126, 393-413.

Cull, W. L., & Zechmeister, E. B. (1994). The learning ability paradox in adult metamemory research: where are the metamemory differences between good and poor learners. Memory and Cognition, 22, 249-257.

Eriksson , Å & Rönnberg, J. (2000). Comprehension calibration and memory prediction accuracy of texts: The effect of time, effort and preference on students' executive monitoring processes. Submitted manuscript.

Gillström, Å. & Rönnberg, J. (1994). Memory prediction accuracy of text recall: Ease, effort and familiarity. Scandinavian Journal of Psychology, 35, 367-386.

Gillström, Å. & Rönnberg, J. (1995). Comprehension calibration and memory prediction accuracy of texts: Reading skill, reading strategies and effort. Journal of Educational Psychology, 87, 545-558.

Glenberg, A. M., & Epstein, W. (1987). Inexpert comprehension calibration of comprehension. Memory and Cognition, 15, 84-93.

Hacker, D. J. (1998). Definitions and empirical foundations. In D. J. Hacker, J. Dunlosky, & Graesser, A. C. (Eds.), Metacognition in educational theory and practice. New Jersey: Lawrence Erlbaum Associates.

Hallam, S., & Francis, H. (1998). Is my understanding yours? A study of higher education students' reading for understanding and the effects of different texts. Learning and Instruction, 8, 83-95.

Koriat, A., & Goldsmith, M. (1997). The role of metacognitive processes in the regulation of memory performance. Chapter to appear in G. Mazzoni & T. O. Nelson' s (Eds.), Monitoring and control processes in metacognition and neuropsychology. IIPDM Report No. 144, October 1997.

Lin L-M., & Zabrucky, K. M. (1998). Calibration of comprehension: research and implications for education and instruction. Contemporary Educational Psychology, 23 345-391.

Logan, G. D. (1988). Toward an instance theory of atomization. Psychological Review, 95, 492-527.

Lovett, S. B., & Pillow, B. H. (1995). Development of the ability to distinguish between comprehension and memory: evidence from strategy-selection tasks. Journal of Educational Psychology, 87, 523-536.

Lovett, S. B., & Pillow, B. H. (1996). Development of the ability to distinguish between comprehension and memory: evidence from goal-state evaluation tasks. Journal of Educational Psychology, 88, 546-562.

Maki, R. H. (1998). Test memory predictions over text material. In D. J. Hacker, Dunlosky, J. & Graesser, A. C. (Eds.), Metacognition in educational theory and practice. New Jersey: Lawrence Erlbaum Associates.

Maki, R., & Berry, S. L. (1984). Metacomprehension of text material. Journal of Experimental Psychology: Learning, Memory, & Cognition, 10, 663-679.

Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances comprehension calibration of comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 609-616.

Maki, R., & Swett, S. (1987). Metamemory for narrative text. Memory and Cognition, 15, 72-83.

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. Journal of Educational Psychology, 88, 49-63.

Mazzoni, G., & Cornoldi, C. (1993). Strategies in study time allocation: Why is study time sometimes not effective? Journal of Experimental Psychology, General, 122, 47-60.

McDaniel, M. A. (1984). The role of elaborative and schema processes of story memory. Memory & Cognition, 12, 46-51.

Necka, E., Machera, M., & Miklas, E. (1992). Incidental learning, intelligence, and verbal ability. Learning and Instruction, 2, 141-153.

Noice, H. (1993). Effects of rote versus gist strategy on the verbatim retention of theatrical scripts. Applied Cognitive Psychology, 7, 75-84.

Pressley, M., & Schneider, W. (1997). Introduction to memory development during childhood and adolescence. London: Lawrence Erlbaum associates.

Pressley, M., Snyder, B. L., Levin, J.R., Murray, H. G., & Ghatala, E. S. (1987). Perceived readiness for examination performance (PREP) produced by initial reading of text and text containing adjunct questions. Reading Research Quarterly, 12, 219-236.

Psykologiförlaget. (1976). Manual till diagnostiska läs- och skrivprov för högstadiet. [Manual for Diagnostic reading-and-writing tests for students grade 7 to 9: Senior level of compulsory school]. Stockholm: Psykologiförlaget.

Schneider, S. L., & Laurion, S. K. (1993). Do we know what we've learned from listening to the news? Memory & Cognition, 21, 198-209.

Stark, R., Renkl, A., Gruber, H., & Mandl, H. (1998). Indeed, sometimes knowledge does not help: A replication study. Instructional Science, 00, 1-17.

Wenestam, C-G. (1993). A critique of research on cognition and cognitive processes. British Journal of Educational Psychology, 63, 34-45.

Westrin, P. A. (1965). WIT III manual. Stockholm: Psykologiförlaget.

Table 1

The experimental design in serial order including the exact formulation of the ratings.

Students' ratings and performances

General ratings of reading and memory abilities.

Fluency: I estimate that my general ability to read texts fluently is (%)?

Comprehension: I estimate that my general ability to comprehend texts is (%)?

Memory: I estimate that my general ability to remember texts is (%)?

Ratings after FREE or FORCED reading

Number of readings: How many times did you read the text through (exact number)?

Prospective ratings of performance

Comprehension calibration: How well have you comprehended the text (in a week, %)?

Memory prediction: How much of the text will you be able to recall (in a week, %)?

Interest: How interesting did you find the text to be (%)?

Performance

Answering performance: 2 multiple-choice questions.

Text recall: as much as you can remember.

Retrospective ratings of performance

Postdiction: How much of the text were you able to recall (in a week, %)?

Postcalibration: How many questions were you able to answer correctly (in a week)?

Evaluation of instructions

Which instruction facilitated recall (REMEMBER, COMPREHEND or BOTH)?

Which instruction facilitated comprehension (REMEMBER, COMPRE., BOTH)?

Open-ended questions about memory and comprehension

Describe how you went through reading for remembering and comprehension.

Name a few things that you think are typical for a good reader.

How do you usually do when you read to understand (+ remember)?

Is it a different or similar task to read to understand and remember?

Table 2

Students' Ratings[*)] of their General Fluency of Reading, Reading Comprehension, Text Memory abilities.

	General Questions % (SD)		
	Fluency	Comprehension	Text Memory
Overall	73.92 (17.54)	71.44 (14.14)	62.75 (16.60)
By verbal skill			
High (n =16)	82.87 (15.11)	76.19 (14.02)	64.94 (14.65)
Medium (n =54)	72.94 (18.48)	70.22 (14.56)	61.78 (17.73)
Low (n =17)	68.89 (14.26)	70.89 (12.74)	63.72 (15.24)

Note: *) ratings ranging from very poor (0%) to very good (100%).

Table 3.

Correlations among Students' General Ratings of Verbal and Memory Abilities and Actual Performances of Text Recall, Level of Comprehension and Verbal Test Scores.

	1	2	3	4	5	6	7		
1. FR	1.00								
2. RC	.44*	1.00							
3. TM	.05	.38*	1.00						
4. Rec-C	.27*	-.08	-.09	1.00					
5. Rec-R	.37*	.04	-.06	.60*	1.00				
6. Com-C	-.03	.18	.17	.27*	.33*	1.00			
7. Com-R	-.06	-.23*	-.09	.30*	.17	.14	1.00		
8. Syn/ant	.24*	.18	.15	.26*	.33*	.12	.20	1.00	
9. Analogy	.17	.11	.02	.13	.23	.12	.05	.51*	1.00

Note: n = 88, * p<.05.

FR, RC, TM: General ratings of fluency of reading, reading comprehension and text memory; Immediate; Rec-C and rec-R: Immediate Text Recall with COMPREHEND and REMEMBER; Com-C and com-R Immediate Level of Comprehension with COMPREHEND and REMEMBER; d) Syn/ant and Analogy: Verbal test scores Synonym-Antonym and Analogy tests.

Table 4

Overall and Verbal skill Immediate and Delayed Mean Comprehension Proportion of Postcalibrations and Level of Comprehension for Both Instructions.

	Immediate testing (n = 88)				Delayed testing (n = 68)			
	COMPREHEND		REMEMBER		COMPREHEND		REMEMBER	
Measure	M	(SD)	M	(SD)	M	(SD)	M	(SD)
Comprehension calibrations %*)								
Overall	67.38	(15.68)	66.77	(18.56)	44.06	(18.22)	40.83	(18.80)
By verbal skill								
High	78.25	(12.53)	74.37	(14.87)	58.31	(10.83)	55.00	(12.30)
Med.	68.05	(15.17)	63.87	(20.50)	40.13	(18.33)	36.15	(19.49)
Low	55.06	(11.38)	68.82	(12.72)	42.56	(17.48)	41.87	(19.32)
Proportion postcalibrations (0-2 questions)								
Overall	.65	(.26)	.65	(.29)	.58	(.30)	.56	(.28)
By verbal skill								
High	.69	(.25)	.76	(.26)	.58	(.19)	.69	(.25)
Med.	.67	(.26)	.64	(.28)	.62	(.31)	.55	(.29)
Low	.54	(.26)	.54	(.33)	.45	(.35)	.45	(.28)
Proportion Level of Comprehension (0-2 questions)								
Overall	.59	(.37)	.66	(.35)	.54	(.38)	.62	(.37)
By verbal skill								
High	.54	(.30)	.54	(.22)	.54	(.41)	.77	(.25)
Med.	.67	(.34)	.44	(.20)	.55	(.35)	.63	(.40)
Low	.37	(.37)	.31	(.20)	.54	(.43)	.42	(.29)

Note: a) Immediate n = 17, delayed n = 13; b) Immediate n = 54, delayed n = 42; c) Immediate n = 17, delayed n = 13. *) ranging from not at all (0%) to very well (100%).

Table 5

Overall and Verbal skill Immediate and Delayed Correlation Coefficients between Comprehension Calibrations and Actual Level of Comprehension and also between Postcalibrations and Actual Level of Comprehension for Both Instructions.

	Immediate (*n*= 88)		Delayed (*n*=68)	
	COMPREHEND	REMEMBER	COMPREHEND	REMEMBER
	Reliabilities of calibration accuracy			
Overall	.14	.05	.20	.05
By verbal skill				
high a)	.27	.46	.24	.24
Med. b)	.18	.03	.18	.05
Low c)	.20	.15	.35	.46
	Reliabilities of postcalibration accuracy			
Overall	.39**	.21*	.19	.19
By verbal skill				
high	.13	.08	.30	.54
Med.	.23	.13	.23	.16
Low	.53	.36	.36	.66*

Note: * = $p < .05$, ** = $p < .01$. a) Immediate *n* = 17, delayed *n* = 13; b) Immediate *n* = 54, delayed *n* = 42; c) Immediate *n* = 17, delayed *n* = 13.

Table 6

Mean Overall and Verbal skill Percentage of Memory Predictions and Memory Postdictions and Proportion of Recall for Both Instructions.

	Immediate testing				Delayed testing			
	COMPREHEND		REMEMBER		COMPREHEND		REMEMBER	
Measure	M	(SD)	M	(SD)	M	(SD)	M	(SD)
				Prediction of text recall %*)				
Overall	56.30	(13.42)	54.85	(14.89)	37.67	(15.35)	35.89	(16.25)
By verbal skill								
High	57.92	(12.45)	60.15	(12.02)	41.85	(12.42)	43.00	(11.02)
Med.	56.73	(14.54)	54.52	(15.61)	37.31	(16.60)	33.44	(17.31)
Low	53.08	(10.60)	50.25	(14.45)	34.33	(14.27)	36.58	(15.96)
				Postdiction of text recall %*)				
Overall	56.05	(18.95)	54.67	(15.99)	28.18	(17.69)	28.89	(18.15)
By verbal skill								
High	67.83	(15.89)	66.42	(11.46)	37.92	(23.31)	36.92	(21.20)
Med.	56.63	(18.06)	53.53	(16.55)	27.93	(15.13)	29.40	(17.21)
Low	43.85	(17.19)	46.46	(12.39)	19.77	(13.60)	20.31	(14.48)
				Mean proportion of text recall				
Overall	.34	(.18)	.40	(.22)	.23	(.13)	.26	(.17)
By verbal skill								
High	.38	(.15)	.50	(.22)	.25	(.12)	.33	(.17)
Med.	.36	(.16)	.41	(.20)	.24	(.14)	.26	(.18)
Low	.25	(.15)	.28	(.20)	.17	(.09)	.16	(.13)

Note: a) Immediate n = 17, delayed n = 13; b) Immediate n = 54, delayed n = 42; c) Immediate n = 17, delayed n = 13. *) ranging from nothing (0%) to everything (100%).

Table 7

Overall and Verbal skill Immediate and Delayed Correlation Coefficients between Memory predictions and Text Recall and also between Postdictions and Text recall for Both Instructions.

	Immediate (n= 88)		Delayed (n=68)	
	COMPREHEND	REMEMBER	COMPREHEND	REMEMBER
	Reliability of prediction accuracy			
Overall	.31**	.29**	.10	.17
By verbal skill				
high a)	-.32	.23	.05	.00
Med. b)	.41**	.36**	.05	.18
Low c)	.20	.52**	.02	.12
	Reliability of postdiction accuracy			
Overall	.44**	.63**	.49**	.53**
By verbal skill				
high	.38	.61*	-.05	-.08
Med.	.41**	.60**	.64**	.60**
Low	.52*	.69*	.51	.60*

Note: * = p < .05, ** = p < .01. a) Immediate n = 17, delayed n = 13; b) Immediate n = 54, delayed n = 42; c) Immediate n = 17, delayed n = 13.

Table 8.

The Table show Mean Recall and Level of Comprehension out of Preference perspective.

	Immediate testing		Delayed testing	
	COMPREHEND	REMEMBER	COMPREHEND	REMEMBER
Measure	M (SD)	M (SD)	M (SD)	M (SD)
Personal best instruction for recall				
Understand[a)]	.40 (.18)	.31 (.18)	.26 (.10)	.22 (.14)
Remember[b)]	.29 (.13)	.48 (.19)	.25 (.11)	.35 (.17)
Both[c)]	.35 (.20)	.40 (.24)	.20 (.14)	.20 (.13)
Personal best instruction for comprehension				
Understand[d)]	.59 (.40)	.65 (.36)	.50 (.38)	.64 (.38)
Remember[e)]	.52 (.33)	.68 (.30)	.46 (.33)	.42 (.36)
Both[f)]	.55 (.41)	.64 (.36)	.56 (.38)	.69 (.38)

a) n = 17, 26, 15, 20 respectively

b) n = 17, 26, 15, 20 respectively

c) n = 35, 35, 26, 26 respectively

d) n = 22, 17, 18, 12 respectively

e) n = 22, 17, 18, 12 respectively

f) n = 40, 40, 31, 32 respectively

Appendix. The Two Experimental Texts Being Used in the Experiment and the Accompanying Comprehension Questions (Psykologiförlaget, 1976). (Correct answer is indicated by asterisk).

Text 1

You probably find the banana a natural product in the supermarket and in the kiosk. In the beginning of the 20th century, however, it was almost an unknown fruit in Europe. The banana was one of the first plants cultivated by the people in East Asia, the original homeland of the banana. When Alexander the Great invaded India in the year 327 B.C., his armies found lots of banana plants by the riverside of the Indus. Presumably, it was then discovered that the dried roots easily could be transported and planted in other hot and humid areas of the world, where good soil was found. The plant immediately put forth new sprouts, spread its large leaves, flourished and bore fruit - - all with amazing speed. Eventually the banana plant dispersed to Africa as well as Australia and the Pacific. A Spanish monk introduced the banana in America shortly after Columbus had discovered the Antilles.

Comprehension Question 1: How is the content of the text best summarized?

a) India - the homeland of the banana.

* b) The banana and its dispersion.

c) The world' s fastest growing tree.

d) The banana' s way to America.

Comprehension Question 2: The banana plant is today found in many different places on earth mainly because

* a) the roots can easily be transported and planted elsewhere.

b) it grows quickly and bears fruit.

c) it can grow anywhere in hot areas.

d) Alexander's armies dispersed it.

e) it can grow anywhere where the soil is good.

Text 2

The sun suddenly rose on the slope where the seven Arab tents were located deep inside the great deserts of Arabia. The tents stood in a half circle, below the sandhills, and in front of every tent rose a blue smoke pillar from a fire of camel dung, where Arabian women with veils thrown back from their faces prepared rice and bread to break the long fast of the night. Around the largest tent sat a group of men on their heels, drawing circles in the sand with their shepherd's sticks or drinking bitter Arabic coffee out of small earless cups. This tent belonged to the leader of the group, and the men in the camp were discussing what they should do. All were barefoot and wore black or brown woolen cloaks over a loose dress of white cloth. Their heads were covered by red and white checked cloths, trimmed with tufts, fixed with black woolen ribbons.

Comprehension Question 1: Which statement is most correct?

a) The Arab women prepared coffee, bread and rice.

b) In front of the largest tent women prepared rice and bread.

c) Women with veils in front of their faces prepared rice.

* d) The men sat around the fire and drank coffee.

e) The men sat by the opening of the tent and drew in the sand.

Comprehension Question 2: Which headline is the best for this paragraph?

* a) Morning in an Arab camp.

b) Sunrise in the desert.

c) Around the fires in the desert.

d) One day among the Arab people.

e) Desert life in Arabia.

LINKÖPING STUDIES IN EDUCATION AND PSYCHOLOGY

1. HOLM, ROLF. Massed and Distributed Criterion Information. Studies of Verbal Learning. 1973. ISBN 91-7222-024-4.

2. TROPÉ, BARBRO. Individual Discovery and External Direction. Studies of the Effects of Different Methods of Instruction. 1974. ISBN 91-7222-078-3.

3. LINDGREN HOOKER, BERNICE. Eductional Flow Models. With Applications to Arab Statistical Data. 1974. ISBN 91-7222-091-0.

4. HOLM, ULLA. Criterion Information and Principle Information. Studies of Verbal Learning. 1975. ISBN 91-7222-108-9.

5. GRUNDIN, HANS. Response Requirements and Information about Correct Responses in Programmed Instruction. 1975. ISBN 91-7372-012-7.

6. ADHAM, KHALIS. Planning Medical Education: Ways to Improve Conditions of Students in Medical Schools in Iraq. 1975. ISBN 91-7372-013-5.

7. ERASMIE, THORD. Language Development and Social Influence. 1975. ISBN 91-7372-038-0.

8. STOCKFELT-HOATSON, BRITT-INGRID. Training of Immigrant Children in Pre-School in Norrköping. 1978. ISBN 91-7372-226-X.

9. CHAIBDERRAINE, MOHAMED. Educational Wastage at the Primary Level. 1978. ISBN 91-7372-221-9.

10. ENGVALL, THORVALD. Förutsättningar och hinder för småföretagsutbildning. 1980. ISBN 91-7372-365-7.

11. BORGBRANT, JAN. Ledningsutveckling. Modellanalys i anslutning till två utvecklingsprojekt. 1980. ISBN 91-7372-366-5.

12. HULT, HÅKAN. Vuxenstuderandes studiesituation. En undersökning utförd inom kommunal vuxenutbildning. 1980. ISBN 91-7372-367-3.

13. HULTMAN, GLENN. Organisationsutveckling genom ledarutbildning. En utvärdering av skolledarutbildningens första utbildningsomgångar. 1981. ISBN 91-7372-452-1.

14. ARNBERG, LENORE. Early Childhood Bilingualism in the Mixed-Lingual Family. Summary. 1981. ISBN 91-7372-462-9.

15. LOTFABADI, HOSSEIN. Higher Education Evaluation: Investigations of Graduate Studies. 1981. ISBN 91-7372-501-3.

16. TÖRNVALL, ANDERS. Läraren och Läroplanen. En undersökning av mellanstadielärares grundsyn i förhållande till mål och riktlinjer i läroplanen för grundskolan (Lgr 69). 1982. ISBN 91-7372-510-2.

17. TÖRNVALL, MARIE-LOUISE. Läraren och Internationaliseringen. En undersökning av mellanstadielärares attityder till internationella frågor i läroplanen för grundskolan (Lgr 69). 1982. ISBN 91-7372-541-2.

18. HOIEN, TORLEIV. Ekkoisk Minne og den Auditive Persepsjonen. 1983. ISBN 91-7372-672-9.

19. ANDRÉ-EKLUND, SIGRID. Vilken frihet ger friheten? En studie av ett deltagarstyrt utvecklingsarbete i gymnasieskolan. 1983. ISBN 91-7372-611-7.

20. AL-ABDI, ABDUL-MUTTALIB. Friluftsverksamheten i innerstadens gymnasieskolor. En undersökning av hur friluftsverksamhetens mål förverkligas i gymnasieskolan i Stockholms innerstad. 1984. ISBN 91-7372-736-9.

21. SVENSSON, TORBJÖRN. Aging and Environment. Institutional Aspects. 1984. ISBN 91-7372-845-4.

22. ELLSTRÖM, PER-ERIK. Rationality, Anarchy, and the Planning of Change in Educational Organizations. A study of Problem Solving and Planning of Change in Small Work Groups. 1984. ISBN 91-7372-852-7.

23. GERHARDSSON, BIRGIT. Kvinnor och vuxenutbildning mitt i livet. En studie av vuxna kvinnors förhållningssätt till utbildning. 1985. ISBN 91-7372-883-7.

24. ANDERSSON, YVONNE. Teknikämnet på grundskolans mellanstadium. 1988. ISBN 91-7870-337-9.

25. WINTER, HANS A. Anställdas deltagande i organisationsutveckling. En teoretisk och empirisk analys av förutsättningarna för participation. 1989. ISBN 91-7870-468-5.

26. NAJI, MOHAMED. En skola för folket - en skola för eliten. En studie av de faktorer som betingar såväl framgång som misslyckanden i Tunisiens primärskolor. 1990. ISBN 91-7870-474-X.

27. MALMBERG, BO. Access to resources in different age-cohorts. Implications for activity level loneliness and life satisfaction. 1990. ISBN 91-7870-592-4.

28. HEDRÉN, ROLF. Logoprogrammering på mellanstadiet. En studie av fördelar och nackdelar med användning av Logo i matematikundervisningen under årskurserna 5 och 6 i grundskolan. 1990. ISBN 91-7870-656-4.

29. BERGSTEN, CHRISTER. Matematisk operativitet. En analys av relationen mellan form och innehåll i skolmatematik. 1990. ISBN 91-7870-666-1.

30. FRANKE, ANITA. Handledning i praktiken. En studie av handledares och lärarkandidaters uppfattningar av handledning i lärarutbildningens praktikdel. 1990. ISBN 91-7870-689-0.

31. FALLSBERG, MARGARETA. Reflections on medicines and medication. A qualitative analysis among people on long-term drug regimens. 1991. ISBN 91-7870-799-4.

32. EKHOLM, BODIL; HEDIN, ANNA. Sitter det i väggarna? En beskrivning av daghemsklimat och barns beteende. 1991. ISBN 91-7870-802-8.

33. SVANBERG HÅRD, HELENE. Informellt lärande. En studie av lärprocesser i folkhögskolemiljö. 1992. ISBN 91-7870-894-X.

34. ERICSON, BRITTA. Editor. Aspects on Literacy. Selected papers from the 13th IRA World Congress on Reading. 1992. ISBN 91-7870-970-9.

35. UNGMARK, INGER. Kvinnor, brott, övervakning. 1992. ISBN 91-7871-012-X.

36. SAMUELSSON, STEFAN. Scripted knowledge packages: Implicit and explicit constraints on comprehension and memory. 1993. ISBN 91-7871-080-4.

37. SJÖSTEN, NILS-ÅKE. Sockenbiblioteket - ett folkbildningsinstrument i 1870-talets Sverige. En studie av folkskoleinspektionens bildningssyn i relation till sockenbiblioteken och den tillgängliga litteraturen. 1993. ISBN 91-7871-093-6.

38. HARTMAN, PER. Skola för ande och hand. En studie av folkhögskolans praktisk-estetiska verksamhet. 1993. ISBN 91-7871-094-4.

39. ELLSTRÖM, EVA. Integration i institutionaliserad verksamhet. En studie av gruppintegration av hörselskadade och döva barn i förskolan. 1993. ISBN 91-7871-118-5.

40. MILSTA, MARGARETHA. Målstyrning och mellanchefers arbete - hierarki, delaktighet och tillfälligheternas spel. 1994. ISBN 91-7871-412-5.

41. PERSSON, ULLA-BRITT. Reading for understanding. An empirical contribution to the metacognition of reading comprehension. 1994. ISBN 91-7871-440-0.

42. REZA-SAHAF, ALI. Language, identity and social behaviour. A sociocultural approach to the study of the concept "will" on the effectiveness of the "how's" and "why's" of bilingualism. 1994. ISBN 91-7871-449-4.

43. AHLSTRAND, ELISABETH. Lärares samarbete - en verksamhet på två arenor. Studier av fyra arbetslag på grundskolans högstadium. 1995. ISBN 91-7871-545-8.

44. COLNERUD, GUNNEL. Etik opch praktik i läraryrket. En empirisk studie av lärares yrkesetiska konflikter i grundskolan. 1995. HLS förlag.

45. RAHIMI, AMENEH. Problem-Based and Conventional Medical Education from a Student Perspective. A qualitative analysis comparing students' experience of medical education, approach to learning and reading comprehension. ISBN 91-7871-584-9.

46. TÖRNQVIST, INGVAR. Oscar Olsssson folkbildaren. I synnerhet hans tankar om universitetens roll i folkbildningsarbetet. 1996. ISBN Sober förlag 91 7296 340-9. ISBN Bildningsförlaget 91 8840 221-5.

47. ANDERSSON, JAN. Two is one too many: Dyadic memory collaboration effects on encoding and retrieval of episodes. 1996. ISBN 91-7871-838-4.

48. SZKLARSKI, ANDRZEJ. Barn och konflikter. En studie av hur konflikter gestaltar sig i svenska och polska barns medvetande. 1996. ISBN 91-7871-842-2.

49. WESTLUND, INGRID. Skolbarn av sin tid. En studie av skolbarns upplevelse av tid. 1996. ISBN 91-7871-847-3.

50. ABRANDT, MADELEINE. Learning Physiotherapy: The impact of formal education and professional experience. 1997. ISBN 91-7871-932-1.

51. SANDAHL, ANITA. Skolmatematiken - kultur eller myt? Mot en bestämning av matematikens didaktiska identitet. 1997. ISBN 91-7871-945-3.

52. EDVARDSSON STIWNE, ELINOR. Förändringsprocesser i kommunal organisation. En studie av organisering och meningsskapande i två förvaltningar. 1997. ISBN 91-7871-944-5.

53. ROBERTSON HÖRBERG, CRISTINA. Lärares kunskapsutnyttjande i praktiken. Ett personligt och kontextuellt perspektiv på vardagskunskap och forskning. 1977. ISBN 91-7219-018-3.

54. TROTZIG, EVA. Sätta flickan i stånd att fullgöra sina husliga plikter. 1997. ISBN 91-7219-034-5.

55. TORSTENSON-ED, TULLIE. Barns livsvägar genom daghem och skola. 1997. ISBN 91-7219-103-1.

56. BERGSTEDT, BO. Den livsupplysande texten. En läsning av N.F.S. Grundt-wigs pedagogiska skrifter. 1998. Carlsson Bokförlag. ISBN-91-7203-281-2.

57. JOHANSSON, KARINA. A "happy" approach to speechreading: The effects of facial expression, emotional content, and script information on speechreading performance. 1998. ISBN-91-7219-228-3.

58. MAGNUSSON, ANDERS. Lärarkunskapens uttryck - en studie av lärares självförståelse och vardagspraktik. 1998. ISBN 91-7219-350-6.

59. ANTONSON, SIVERT. Hörselskadade i högskolestudier. Möjligheter och Hinder. 1998. ISBN 91-7219-360-3.

60. SALMINEN-KARLSSON, MINNA. Bringing Women into Computer Engineering. Curriculum Reform Processes at Two Institutes of Technology. 1999. ISBN 91-7219-484-7.

61. OSMAN, ALI. The Strangers among us. Ethnicity, Culture, and Immigrants: The Social Construction of Identity in Adult Education. 1999. ISBN 91-7219-530-4.

62. ADAMSON, LENA. Like Circles on the Water. A Study of Adolescent Identity. 1999. ISBN 91-7219-533-9.

63. DAVIDSON, BO; SVEDIN, PER-OLOF. Lärande i produktionssystem. - En studie av operatörsarbete i högautomatiserad process- och verkstadsindustri. 1999. ISBN 91-7219-535-5.

64 THUNBORG, CAMILLA. Lärande av yrkesidentiteter. – En studie av läkare, sjuksköterskor och undersköterskor. 1999. ISBN 91-7219-555-x

65, GÖRANSSON, ANNE. Vision in sight. The relationships between knowledge, health beliefs and treatment outcomes. The case of amblyopia. 1999. ISBN 91-7219-621-1.

66. LANTZ, AGNETA. Breaking Information Barriers Through Information Literacy. A Longitudinal and Interventional Study among Small-Firm Managers. 1999. ISBN 91-7219-622-x.

67. SCHOULTZ, JAN. Att samtala om/i naturvetenskap. Kommunikation, kontext och artefakt. 2000. ISBN 91-7219-653-x.

68. ANDERSSON, PER. Att studera och bli bedömd. Empiriska och teoretiska perspektiv på gymnasie- och vuxenstuderandes sätt att erfara studier och bedömningar. 2000. ISBN 91-7219-675-0.

69. JALONEN, HARRI. Flygutbildning och flygsäkerhet. En studie av relationen mellan utbildningskarriär och haveribenägenhet i ett Human Factors-perspektiv samt en attitydundersökning. 2000. ISBN 91-7219-710-2.